



THE  
POWER  
TO KNOW.

# **SAS/STAT<sup>®</sup> 9.22 User's Guide**

## **The LIFETEST Procedure**

### **(Book Excerpt)**



This document is an individual chapter from *SAS/STAT® 9.22 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2010. *SAS/STAT® 9.22 User's Guide*. Cary, NC: SAS Institute Inc.

Copyright © 2010, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

**For a Web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

**U.S. Government Restricted Rights Notice:** Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st electronic book, May 2010

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at [support.sas.com/publishing](http://support.sas.com/publishing) or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

## Chapter 49

# The LIFETEST Procedure

### Contents

Overview: LIFETEST Procedure . . . . .	<b>3703</b>
Getting Started: LIFETEST Procedure . . . . .	<b>3705</b>
Syntax: LIFETEST Procedure . . . . .	<b>3714</b>
PROC LIFETEST Statement . . . . .	3714
BY Statement . . . . .	3728
FREQ Statement . . . . .	3729
ID Statement . . . . .	3729
STRATA Statement . . . . .	3730
TEST Statement . . . . .	3735
TIME Statement . . . . .	3735
Details: LIFETEST Procedure . . . . .	<b>3735</b>
Missing Values . . . . .	3735
Computational Formulas . . . . .	3736
Computer Resources . . . . .	3752
Output Data Sets . . . . .	3753
OUTSURV= Data Set . . . . .	3753
OUTTEST= Data Set . . . . .	3754
Displayed Output . . . . .	3755
ODS Table Names . . . . .	3762
ODS Graphics . . . . .	3764
Examples: LIFETEST Procedure . . . . .	<b>3766</b>
Example 49.1: Product-Limit Estimates and Tests of Association . . . . .	3766
Example 49.2: Enhanced Survival Plot and Multiple-Comparison Adjustments . . . . .	3779
Example 49.3: Life-Table Estimates for Males with Angina Pectoris . . . . .	3783
References . . . . .	<b>3790</b>

---

## Overview: LIFETEST Procedure

A common feature of lifetime or survival data is the presence of right-censored observations due either to withdrawal of experimental units or to termination of the experiment. For such observations, you know only that the lifetime exceeded a given value; the exact lifetime remains unknown. Such

data cannot be analyzed by ignoring the censored observations because, among other considerations, the longer-lived units are generally more likely to be censored. The analysis methodology must correctly use the censored observations as well as the uncensored observations.

Texts that discuss the survival analysis methodology include Collett (1994), Cox and Oakes (1984), Kalbfleisch and Prentice (1980), Klein and Moeschberger (1997), Lawless (1982), and Lee (1992). Users interested in the theory should consult Fleming and Harrington (1991) and Andersen et al. (1992).

Usually, a first step in the analysis of survival data is the estimation of the distribution of the survival times. Survival times are often called *failure* times, and *event* times are uncensored survival times. The survival distribution function (SDF), also known as the survivor function, is used to describe the lifetimes of the population of interest. The SDF evaluated at  $t$  is the probability that an experimental unit from the population will have a lifetime exceeding  $t$ —that is,

$$S(t) = \Pr(T > t)$$

where  $S(t)$  denotes the survivor function and  $T$  is the lifetime of a randomly selected experimental unit. The LIFETEST procedure can be used to compute nonparametric estimates of the survivor function either by the product-limit method (also called the Kaplan-Meier method) or by the life-table method (also called the actuarial method). The life-table estimator is a grouped-data analog of the Kaplan-Meier estimator. The procedure can also compute the Breslow estimator or the Fleming-Harrington estimator, which are asymptotic equivalent alternatives to the Kaplan-Meier estimator.

Some functions closely related to the SDF are the cumulative distribution function (CDF), the probability density function (PDF), and the hazard function. The CDF, denoted  $F(t)$ , is defined as  $1 - S(t)$  and is the probability that a lifetime does not exceed  $t$ . The PDF, denoted  $f(t)$ , is defined as the derivative of  $F(t)$ , and the hazard function, denoted  $h(t)$ , is defined as  $f(t)/S(t)$ . If the life-table method is chosen, the estimates of the probability density function can also be computed. Plots of these estimates can be produced by a graphical or line printer device, or based on the output delivery system (ODS).

An important task in the analysis of survival data is the comparison of survival curves. It is of interest to determine whether the underlying populations of  $k$  ( $k \geq 2$ ) samples have identical survivor functions. PROC LIFETEST provides nonparametric  $k$ -sample tests based on weighted comparisons of the estimated hazard rate of the individual population under the null and alternative hypotheses. Corresponding to various weight functions, a variety of tests can be specified, which include the log-rank test, Wilcoxon test, Tarone-Ware test, Peto-Peto test, modified Peto-Peto test, and Fleming-Harrington  $G_\rho$  family of tests. PROC LIFETEST also provides corresponding trend tests to detect ordered alternatives. Stratified tests can be specified to adjust for prognostic factors that affect the events rates in the various populations. A likelihood ratio test, based on an underlying exponential model, is also included to compare the survival curves of the samples.

There are other prognostic variables, called covariates, that are thought to be related to the failure time. These covariates can also be used to construct statistics to test for association between the covariates and the lifetime variable. PROC LIFETEST can compute two such test statistics: censored data



linear rank statistics based on the exponential scores and the Wilcoxon scores. The corresponding tests are known as the log-rank test and the Wilcoxon test, respectively. These tests are computed by pooling over any defined strata, thus adjusting for the stratum variables.

One change in SAS 9.2 is that the calculation of confidence limits for the quartiles of survival time is based on the transformation specified by the `CONFTYPE=` option. Another change is that the `SURVIVAL` statement in SAS 9.1 is folded into the `PROC LIFETEST` statement; that is, options that were in the `SURVIVAL` statement can now be specified in the `PROC LIFETEST` statement. The `SURVIVAL` statement is no longer needed and it is not documented.

---

## Getting Started: LIFETEST Procedure

You can use the `LIFETEST` procedure to compute nonparametric estimates of the survivor functions, to compare survival curves, and to compute rank tests for association of the failure time variable with covariates.

For simple analyses, only the `PROC LIFETEST` and `TIME` statements are required. Consider a sample of survival data. Suppose that the time variable is `T` and the censoring variable is `C` with value 1 indicating censored observations. The following statements compute the product-limit estimate for the sample:

```
proc lifetest;
  time t*c(1);
run;
```

You can use the `STRATA` statement to divide the data into various strata. A separate survivor function is then estimated for each stratum, and tests of the homogeneity of strata are performed. However, if the `GROUP=` option is also specified in the `STRATA` statement, the `GROUP=` variable is used to identify the samples whose survivor functions are to be compared, and the `STRATA` variables are used to define the strata for the stratified tests. You can specify covariates (prognostic variables) in the `TEST` statement, and `PROC LIFETEST` computes linear rank statistics to test the effects of these covariates on survival.

For example, consider the results of a small randomized trial on rats. Suppose you randomize 40 rats that have been exposed to a carcinogen into two treatment groups (Drug X and Placebo). The event of interest is death from cancer induced by the carcinogen. The response is the time from randomization to death. Four rats died of other causes; their survival times are regarded as censored observations. Interest lies in whether the survival distributions differ between the two treatments.

The following `DATA` step creates the data set `Exposed`, which contains four variables: `Days` (survival time in days from treatment to death), `Status` (censoring indicator variable: 0 if censored and 1 if not censored), `Treatment` (treatment indicator), and `Sex` (gender: F if female and M if male).

```

proc format;
  value Rx 1='Drug X' 0='Placebo';
data exposed;
  input Days  Status Treatment Sex $ @@;
  format Treatment Rx.;
  datalines;
179  1  1  F  378  0  1  M
256  1  1  F  355  1  1  M
262  1  1  M  319  1  1  M
256  1  1  F  256  1  1  M
255  1  1  M  171  1  1  F
224  0  1  F  325  1  1  M
225  1  1  F  325  1  1  M
287  1  1  M  217  1  1  F
319  1  1  M  255  1  1  F
264  1  1  M  256  1  1  F
237  0  0  F  291  1  0  M
156  1  0  F  323  1  0  M
270  1  0  M  253  1  0  M
257  1  0  M  206  1  0  F
242  1  0  M  206  1  0  F
157  1  0  F  237  1  0  M
249  1  0  M  211  1  0  F
180  1  0  F  229  1  0  F
226  1  0  F  234  1  0  F
268  0  0  M  209  1  0  F
;

```

PROC LIFETEST is invoked as follows to compute the product-limit estimate of the survivor function for each treatment and to compare the survivor functions between the two treatments.

```

ods graphics on;
proc lifetest data=Exposed plots=(survival(atrisk) logsurv);
  time Days*Status(0);
  strata Treatment;
run;
ods graphics off;

```

In the TIME statement, the survival time variable, Days, is crossed with the censoring variable, Status, with the value 0 indicating censoring. That is, the values of Days are considered censored if the corresponding values of Status are 0; otherwise, they are considered as event times. In the STRATA statement, the variable Treatment is specified, which indicates that the data are to be divided into strata based on the values of Treatment. Graphics results are enabled through ODS with the specification of the ODS GRAPHICS ON statement. Two plots are requested through the PLOTS= option—a plot of the survival curves with at risk numbers and a plot of the negative log of the survival curves.

The results of the analysis are displayed in the following figures.

Figure 49.1 displays the product-limit survival estimate for the Drug X group (Treatment=1). The figure lists, for each observed time, the survival estimate, failure rate, standard error of the estimate, cumulative number of failures, and number of subjects remaining in the study.

**Figure 49.1** Survivor Function Estimate for the Drug X-Treated Rats

The LIFETEST Procedure					
Stratum 1: Treatment = Drug X					
Product-Limit Survival Estimates					
Days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	20
171.000	0.9500	0.0500	0.0487	1	19
179.000	0.9000	0.1000	0.0671	2	18
217.000	0.8500	0.1500	0.0798	3	17
224.000*	.	.	.	3	16
225.000	0.7969	0.2031	0.0908	4	15
255.000	.	.	.	5	14
255.000	0.6906	0.3094	0.1053	6	13
256.000	.	.	.	7	12
256.000	.	.	.	8	11
256.000	.	.	.	9	10
256.000	0.4781	0.5219	0.1146	10	9
262.000	0.4250	0.5750	0.1135	11	8
264.000	0.3719	0.6281	0.1111	12	7
287.000	0.3187	0.6813	0.1071	13	6
319.000	.	.	.	14	5
319.000	0.2125	0.7875	0.0942	15	4
325.000	.	.	.	16	3
325.000	0.1062	0.8938	0.0710	17	2
355.000	0.0531	0.9469	0.0517	18	1
378.000*	0.0531	.	.	18	0

NOTE: The marked survival times are censored observations.

Figure 49.2 displays summary statistics of survival times for the Drug X group. It contains estimates of the 25th, 50th, and 75th percentiles and the corresponding 95% confidence limits. The median survival time for rats in this treatment is 256 days. The mean and standard error are also displayed; however, it is noted that these values are underestimated because the largest observed time is censored and the estimation is restricted to the largest event time.

**Figure 49.2** Summary Statistics of Survival Times for Drug X-Treated Rats

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	319.000	LOGLOG	256.000	355.000
50	256.000	LOGLOG	255.000	319.000
25	255.000	LOGLOG	171.000	256.000

**Figure 49.2** *continued*

Mean	Standard Error
271.131	11.877
NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.	

Figure 49.3 and Figure 49.4 display the survival estimates and the summary statistics of the survival times for Placebo (Treatment=0). The median survival time for rats in this treatment is 235 days.

**Figure 49.3** Survivor Function Estimate for Placebo-Treated Rats

The LIFETEST Procedure					
Stratum 2: Treatment = Placebo					
Product-Limit Survival Estimates					
Days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	20
156.000	0.9500	0.0500	0.0487	1	19
157.000	0.9000	0.1000	0.0671	2	18
180.000	0.8500	0.1500	0.0798	3	17
206.000	.	.	.	4	16
206.000	0.7500	0.2500	0.0968	5	15
209.000	0.7000	0.3000	0.1025	6	14
211.000	0.6500	0.3500	0.1067	7	13
226.000	0.6000	0.4000	0.1095	8	12
229.000	0.5500	0.4500	0.1112	9	11
234.000	0.5000	0.5000	0.1118	10	10
237.000	0.4500	0.5500	0.1112	11	9
237.000*	.	.	.	11	8
242.000	0.3938	0.6063	0.1106	12	7
249.000	0.3375	0.6625	0.1082	13	6
253.000	0.2813	0.7188	0.1038	14	5
257.000	0.2250	0.7750	0.0971	15	4
268.000*	.	.	.	15	3
270.000	0.1500	0.8500	0.0891	16	2
291.000	0.0750	0.9250	0.0693	17	1
323.000	0	1.0000	.	18	0
NOTE: The marked survival times are censored observations.					

**Figure 49.4** Summary Statistics of Survival Times for Placebo-Treated Rats

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	257.000	LOGLOG	237.000	323.000
50	235.500	LOGLOG	206.000	253.000
25	207.500	LOGLOG	156.000	229.000
Mean Standard Error				
235.156		10.211		

A summary of the number of censored and event observations is shown in [Figure 49.5](#). The figure lists, for each stratum, the number of event and censored observations, and the percentage of censored observations.

**Figure 49.5** Number of Event and Censored Observations

Summary of the Number of Censored and Uncensored Values					
Stratum	Treatment	Total	Failed	Censored	Percent Censored
1	Drug X	20	18	2	10.00
2	Placebo	20	18	2	10.00
<hr/>					
Total		40	36	4	10.00

[Figure 49.6](#) displays the graph of the product-limit survivor function estimates versus survival time. The two treatments differ primarily at larger survival times. Note the number of subjects at risk in the plot. You can display the number of subjects at risk at specific time points by using the [ATRISK=](#) option.

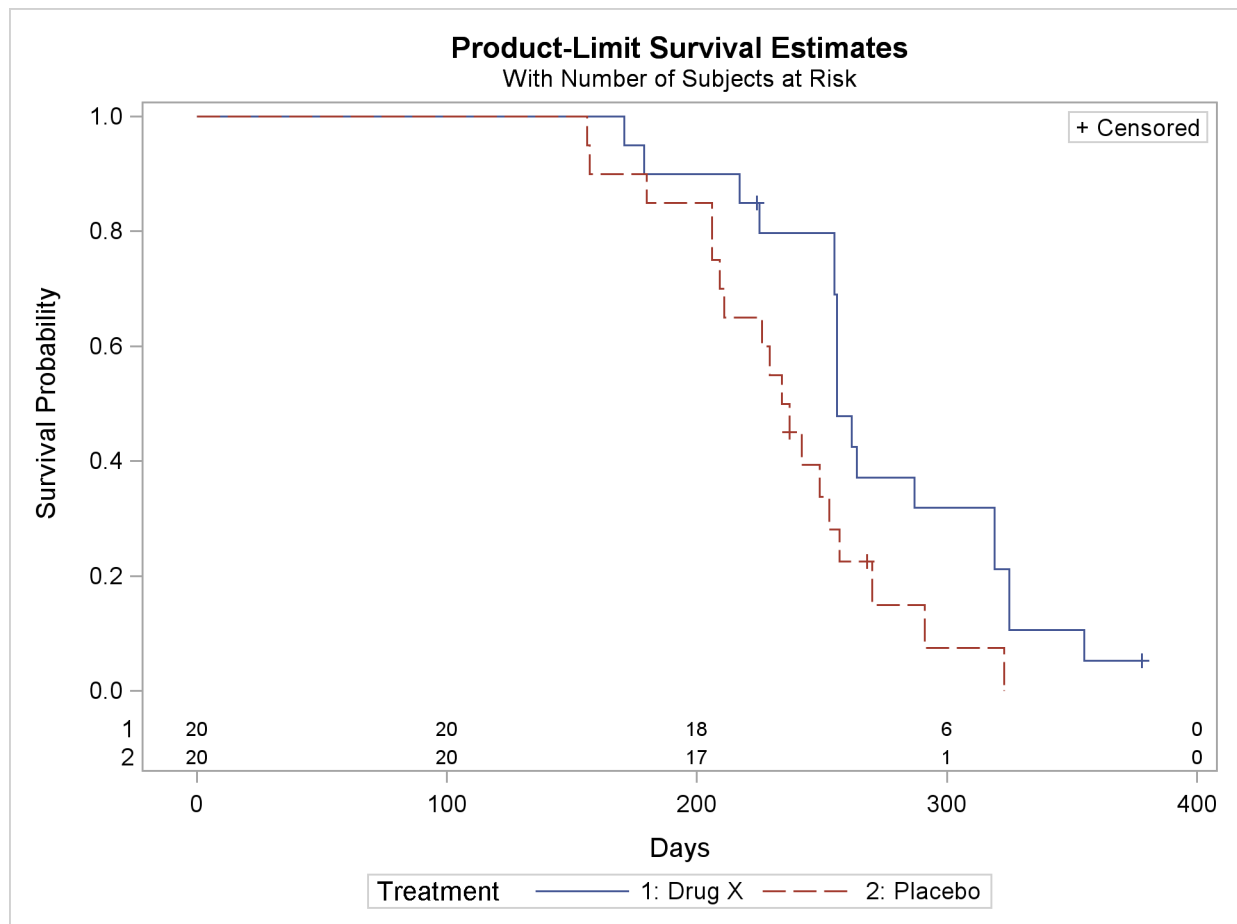
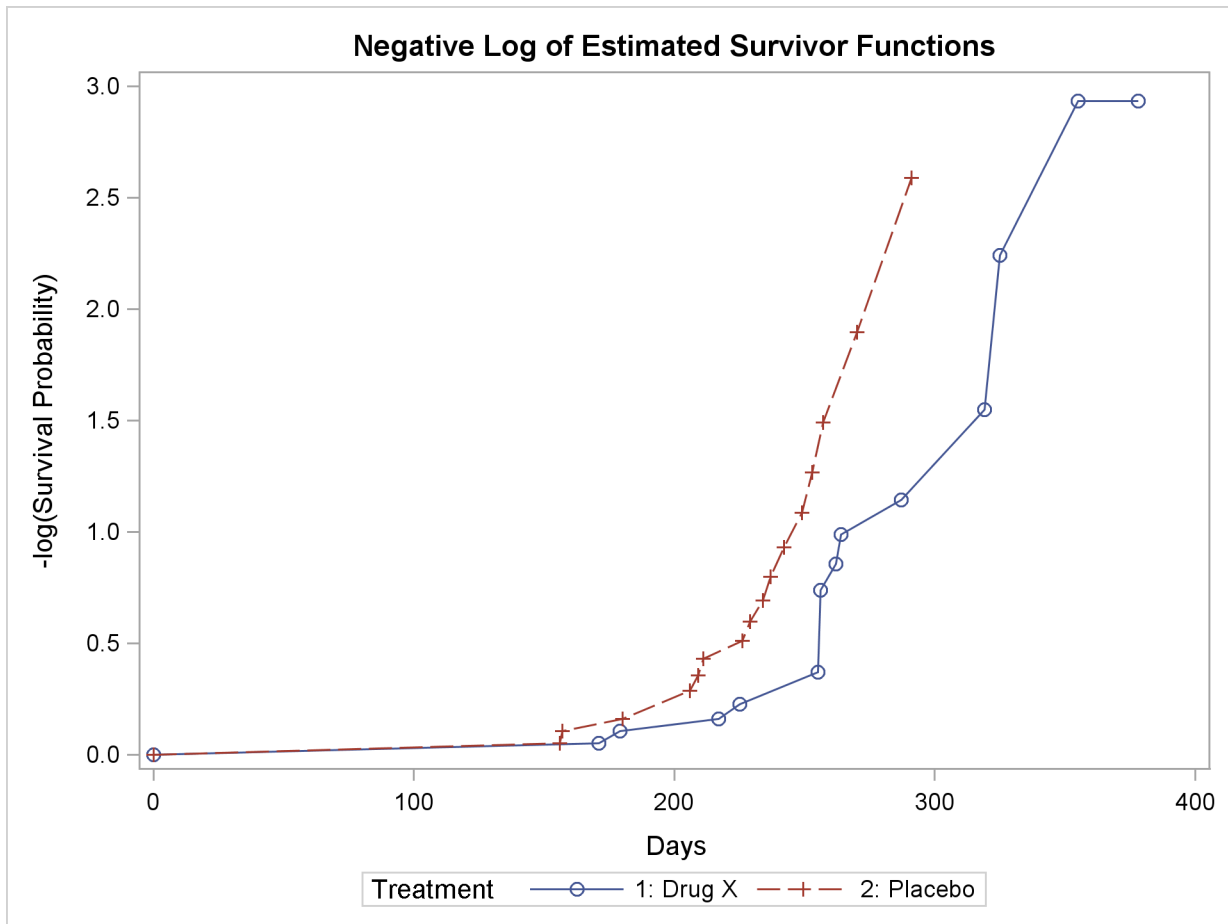
**Figure 49.6** Plot of Estimated Survivor Functions

Figure 49.7 displays the graph of the log survivor function estimates versus survival time. Neither curve approximates a straight line through the origin—the exponential model is not appropriate for the survival data.

Note that these graphical displays are generated through ODS. For general information about ODS Graphics, see Chapter 21, “[Statistical Graphics Using ODS](#).”

**Figure 49.7** Plot of Estimated Negative Log Survivor Functions

Results of the comparison of survival curves between the two treatments are shown in [Figure 49.8](#). The rank tests for homogeneity indicate a significant difference between the treatments ( $p=0.0175$  for the log-rank test and  $p=0.0249$  for the Wilcoxon test). Rats treated with Drug X live significantly longer than those treated with Placebo. Since the survival curves for the two treatments differ primarily at longer survival times, the Wilcoxon test, which places more weight on shorter survival times, becomes less significant than the log-rank test. As noted earlier, the exponential model is not appropriate for the given survival data; consequently, the result of the likelihood ratio test should be ignored.

**Figure 49.8** Results of the Two-sample Tests

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.6485	1	0.0175
Wilcoxon	5.0312	1	0.0249
-2Log(LR)	0.1983	1	0.6561

Next, suppose male rats and female rats are thought to have different survival rates, and you want to assess the treatment effect while adjusting for the gender differences. By specifying the variable `Sex` in the `STRATA` statement as a stratifying variable and by specifying the variable `Treatment` in the `GROUP=` option, you can carry out a stratified test to test `Treatment` while adjusting for `Sex`. The test statistics are computed by pooling over the strata defined by the values of `Sex`, thus controlling for the effect of `Sex`. The `NOTABLE` option is added to the `PROC LIFETEST` statement as follows to avoid estimating a survival curve for each gender.

```
proc lifetest data=Exposed notable;
  time Days*Status(0);
  strata Sex / group=Treatment;
run;
```

Results of the stratified tests are shown in Figure 49.9. The treatment effect is statistically significant for both the log-rank test ( $p=0.0071$ ) and the Wilcoxon test ( $p=0.0150$ ). As compared to the results of the unstratified tests in Figure 49.8, the significance of the treatment effect has been sharpened by controlling for the effect of the gender of the subjects.

**Figure 49.9** Results of the Stratified Two-Sample Tests

The LIFETEST Procedure			
Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	7.2466	1	0.0071
Wilcoxon	5.9179	1	0.0150

Since `Treatment` is a binary variable, another way to study the effect of `Treatment` is to carry out a censored linear rank test with `Treatment` as an independent variable. Although this test is less popular than the two-sample test, nevertheless, in situations where the independent variables are continuous and are difficult to discretize, it might be infeasible to perform a  $k$ -sample test. To compute the censored linear rank statistics to test the `Treatment` effect, `Treatment` is specified in the `TEST` statement as follows.

```
proc lifetest data=Exposed notable;
  time Days*Status(0);
  test Treatment;
run;
```

Results of the linear rank tests are shown Figure 49.10. The  $p$ -values are very similar to those of the two-sample tests in Figure 49.8.



**Figure 49.10** Results of Linear Rank Tests of Treatment

The LIFETEST Procedure				
Univariate Chi-Squares for the Wilcoxon Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
Treatment	3.9525	1.7524	5.0875	0.0241
Univariate Chi-Squares for the Log-Rank Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
Treatment	6.2708	2.6793	5.4779	0.0193

With Sex as a prognostic factor that you want to control, you can compute a stratified linear rank statistic to test the effect of Treatment by specifying Sex in the STRATA statement and Treatment in the TEST statement as in the following program. The TEST=NONE option is specified in the STRATA statement to suppress the two-sample tests for Sex.

```
proc lifetest data=Exposed notable;
  time Days*Status(0);
  strata Sex / test=none;
  test Treatment;
run;
```

Results of the stratified linear rank tests are shown in [Figure 49.11](#). The *p*-values are very similar to those of the stratified tests in [Figure 49.9](#).

**Figure 49.11** Results of Stratified Linear Rank Tests of Treatment

The LIFETEST Procedure				
Univariate Chi-Squares for the Wilcoxon Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
Treatment	4.2372	1.7371	5.9503	0.0147
Univariate Chi-Squares for the Log-Rank Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
Treatment	6.8021	2.5419	7.1609	0.0075

---

## Syntax: LIFETEST Procedure

The following statements are available in PROC LIFETEST:

```
PROC LIFETEST < options > ;
    BY variables ;
    FREQ variable ;
    ID variables ;
    STRATA variable < (list) > < ... variable < (list) > > < /options > ;
    TEST variables ;
    TIME variable < *censor(list) > ;
```

The simplest use of PROC LIFETEST is to request the nonparametric estimates of the survivor function for a sample of survival times. In such a case, only the PROC LIFETEST statement and the TIME statement are required. You can use the STRATA statement to divide the data into various strata. A separate survivor function is then estimated for each stratum, and tests of the homogeneity of strata are performed. However, if the GROUP= option is also specified in the STRATA statement, stratified tests are carried out to test the  $k$  samples defined by the GROUP= variable while controlling for the effect of the STRATA variables. You can specify covariates in the TEST statement. PROC LIFETEST computes linear rank statistics to test the effects of these covariates on survival.

The PROC LIFETEST statement invokes the procedure. All statements except the TIME statement are optional, and there is no required order for the statements following the PROC LIFETEST statement. The TIME statement is used to specify the variables that define the survival time and censoring indicator. The STRATA statement specifies a variable or set of variables defining the strata for the analysis. The TEST statement specifies a list of numeric covariates to be tested for their association with the response survival time. Each variable is tested individually, and a joint test statistic is also computed. The ID statement provides a list of variables whose values are used to identify observations in the product-limit, Breslow, or Fleming-Harrington estimates. When only the TIME statement appears, no strata are defined and no tests of homogeneity are performed.

---

## PROC LIFETEST Statement

```
PROC LIFETEST < options > ;
```

The PROC LIFETEST statement invokes the procedure. Optionally, this statement identifies an input and an OUTSURV= data set, and specifies the computation details of the survivor function estimation. The options listed in [Table 49.1](#) are available in the PROC LIFETEST statement and are described in alphabetic order. If no options are requested, PROC LIFETEST computes and displays the product-limit estimate of the survivor function; and if an ODS GRAPHICS ON statement is specified, a plot of the estimated survivor function is also displayed.

**Table 49.1** Options Available in the PROC LIFETEST Statement

Option	Description
<b>Input and Output Data Sets</b>	
DATA=	Specifies the input SAS data set
OUTSURV=	Names an output data set to contain survival estimates and confidence limits
OUTTEST=	Names an output data set to contain rank test statistics for association of survival time with covariates
<b>Nonparametric Estimation</b>	
INTERVALS=	Specifies interval endpoints for life-table estimates
NELSON	Adds the Nelson-Aalen estimates
METHOD=	Specifies the method to compute survivor function
NINTERVAL=	Specifies the number of intervals for life-table estimates
WIDTH=	Specifies the width of intervals for life-table estimates
<b>Confidence Limits for Survivorship</b>	
ALPHA=	Sets the confidence level for interval estimation estimates
BANDMAXTIME=	Specifies the maximum time for confidence band
BANDMINTIME=	Specifies the minimum time for confidence band
CONFBAND=	Specifies the type of confidence band in the OUTSURV= data set
CONFTYPE=	Specifies the transformation applied to the survivor function to obtain confidence limits
<b>Line Printer Plots</b>	
FORMCHAR(1,2,7,9)=	Defines the characters used for line printer plot axes
LINEPRINTER	Specifies that plots be produced by a line printer
MAXTIME=	Specifies the maximum time value for plotting
NOCENSLOT	Suppresses the plot of censored observations
PLOTS=	Specifies the plots to display
<b>ODS Graphics</b>	
MAXTIME=	Specifies the maximum time value for plotting
PLOTS=	specifies plots to display
<b>Traditional Graphics</b>	
ANNOTATE=	Specifies an Annotate data set that adds features to plots
CENSORED SYMBOL=	Defines the symbol used for censored observations in plots
DESCRIPTION=	specifies the string that appears in the description field of the PROC GREPLAY master menu for the plots
EVENTSYMBOL=	Specifies the symbol used for event observations in plots
GOUT=	Specifies the graphics catalog name for saving graphics output
LANNOTATE=	Specifies an input data set that contains variables for local annotation
MAXTIME=	Specifies the maximum time value for plotting
PLOTS=	Specifies the plots to display
<b>Control Output</b>	
ATRISK	Adds the number of subjects at risk to the survival estimate table
NO PRINT	Suppresses the display of printed output

**Table 49.1** *continued*

Option	Description
NOTABLE	Suppresses the display of survival function estimates
INTERVALS=	Displays only the estimate for the smallest time in each interval
NOLEFT	Suppresses the Number Left column in the survival estimate table
TIMELIST=	Specifies a list of time points to display the survival estimate
REDUCEOUT	Specifies that only INTERVAL= or TIMELIST= observations be listed in the OUTSURV= data set
<b>Miscellaneous</b>	
ALPHAQT=	Sets the confidence level for survival time quartiles
MISSING	Allows missing values to be a stratum level
SINGULAR=	Sets the tolerance for testing singularity of covariance matrix of rank statistics
STDERR	Outputs the standard error for the survival estimators to the OUTSURV= data set
TIMELIM=	Specifies the time limit used to estimate the mean survival time and its standard error

The PLOTS= option in the PROC LIFETEST statement specifies the plots to display. You can select one of the following three types of graphics in PROC LIFETEST: line printer, traditional, and ODS. If you specify the **LINEPRINTER** option, you get line printer plots; otherwise you get the traditional graphics if ODS Graphics is not enabled, and you get the ODS Graphics plots if the ODS GRAPHICS ON statement is specified. The following table depicts the behavior regarding the type of graphics produced:

ODS Graphics	PROC LIFETEST Options		Behavior
OFF			No graphics
OFF		LINEPRINTER	No graphics
OFF	PLOTS=SURVIVAL		Traditional graphics
OFF	PLOTS=SURVIVAL	LINEPRINTER	Line printer plot
ON			ODS Graphics survival plot
ON		LINEPRINTER	No graphics
ON	PLOTS=SURVIVAL		ODS Graphics
ON	PLOTS=SURVIVAL	LINEPRINTER	Line printer plot

ODS Graphics is now the preferred method of requesting graphics. Many new features have been added to the ODS Graphics plots in PROC LIFETEST. For example, you can display the number of subjects at risk in a survival plot through ODS Graphics, but such a feature is not available in traditional graphics or line printer plots. The PLOTS= option syntax is documented separately for each type of graphics and is preceded by a heading that indicates the graphics type.

**ALPHA= $\alpha$** 

specifies the level of significance  $\alpha$  for the  $100(1 - \alpha)\%$  confidence intervals for the survivor, hazard, and density functions. For example, the option ALPHA=0.05 requests the 95% confidence limits for the survivor function. The default value is 0.05.

**ALPHAQT= $\alpha$** 

specifies the significance level  $\alpha$  for the  $100(1 - \alpha)\%$  confidence intervals for the quartiles of the survival time. For example, the option ALPHAQT=0.05 requests a 95% confidence interval for the quartiles of the survival time. The default value is 0.05.

**ANNOTATE=SAS-data-set****ANNO=SAS-data-set**

specifies an input data set that contains appropriate variables for annotation of the traditional graphics. The ANNOTATE= option enables you to add features (for example, labels explaining extreme observations) to plots produced on graphics devices. The ANNOTATE= option cannot be used if the [LINEPRINTER](#) option or the ODS GRAPHICS ON statement is specified. The data set specified must be an ANNOTATE= type data set, as described in *SAS/GRAPH Software: Reference*.

The data set specified with the ANNOTATE= option in the PROC LIFETEST statement is “global” in the sense that the information in this data set is displayed in every plot produced by a single invocation of PROC LIFETEST.

**ATRISK**

adds a column that represents the number of subjects at risk to the survival estimate table. Also added is a column that represents the number of events at each observed time. This option has no effect for the life-table method.

**BANDMAXTIME=value****BANDMAX=value**

specifies the maximum time for the confidence bands. The default is the largest observed event time. If the specified BANDMAX= time exceeds the largest observed event time, it is truncated to the largest observed event time.

**BANDMINTIME=value****BANDMIN=value**

specifies the minimum time for the confidence bands. The default is the smallest observed event time. For the equal-precision band, if the BANDMIN= value is less than the smallest observed event time, it is defaulted to the smallest observed event time.

**CENSORED SYMBOL=name | 'string'****CS=name | 'string'**

specifies the symbol value for the censored observations in traditional graphics. The value, *name* or *'string'*, is the symbol value specification allowed in SAS/GRAPH software. The default is CS=CIRCLE. If you want to omit plotting the censored observations, specify CS=NONE. The CENSORED SYMBOL= option cannot be used if the [LINEPRINTER](#) option or the ODS GRAPHICS ON statement is specified.

**CONFBAND=keyword**

specifies the confidence bands to be output to the OUTSURV= data set. Confidence bands are available for METHOD=KM, METHOD=BRESLOW, or METHOD=FH. You can use the following *keywords*:

ALL                      outputs both the Hall-Wellner and the equal-precision confidence bands.

EP	outputs the equal-precision confidence bands.
HW	outputs the Hall-Wellner confidence bands.

**CONFTYPE=***keyword*

specifies the transformation applied to  $S(t)$  to obtain the pointwise confidence intervals and the confidence bands for the survivor function as well as the confidence intervals for the quartiles of the survival times. The following *keywords* can be used and the default is CONFTYPE=LOGLOG.

ASINSQRT	the arcsine-square root transformation
----------	--

$$g(x) = \sin^{-1}(\sqrt{x})$$

LOGLOG	the log-log transformation
--------	----------------------------

$$g(x) = \log(-\log(x))$$

This is also referred to as the log cumulative hazard transformation since it is applying the logarithmic function  $\log(\cdot)$  to the cumulative hazard function. Collett (1994) and Lachin (2000) refer it as the complementary log-log transformation.

LINEAR	the identity transformation
--------	-----------------------------

$$g(x) = x$$

LOG	the logarithmic transformation
-----	--------------------------------

$$g(x) = \log(x)$$

LOGIT	the logit transformation
-------	--------------------------

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

**DATA=***SAS-data-set*

names the SAS data set used by PROC LIFETEST. By default, the most recently created SAS data set is used.

**DESCRIPTION=***'string'***DES=***'string'*

specifies a descriptive string of up to 256 characters that appears in the “Description” field of the traditional graphics catalog. The description does not appear in the plots. By default, PROC LIFETEST assigns a description of the form PLOT OF *vname* vs *hname*, where *vname* and *hname* are the names of the *y* variable and the *x* variable, respectively. The DESCRIPTION= option cannot be used if the **LINEPRINTER** option or the ODS GRAPHICS ON statement is specified.

**EVENTSYMBOL**=*name* | '*string*'

**ES**=*name* | '*string*'

specifies the symbol value for the event observations in traditional graphics. The value, *name* or '*string*', is the symbol value specification allowed in SAS/GRAPH software. The default is ES=NONE. The EVENTSYMBOL= option cannot be used if the [LINEPRINTER](#) option or the ODS GRAPHICS ON statement is specified.

**FORMCHAR**(1,2,7,9)='*string*'

defines the characters used for constructing the vertical and horizontal axes of the line printer plots. The string should be four characters. The first and second characters define the vertical and horizontal bars, respectively, which are also used in drawing the *steps* of the Kaplan-Meier, Breslow, or Fleming-Harrington survival curve. The third character defines the tick mark for the axes, and the fourth character defines the lower left corner of the plot. The default is FORMCHAR(1,2,7,9)='|-+-'. Any character or hexadecimal string can be used to customize the plot appearance. If you use hexadecimal, you must put an x after the closing quote. For example, to send the plot output to a printer with the IBM graphics character set (1 or 2), specify

```
formchar (1, 2, 7, 9) = 'B3C4C5C0' x
```

Refer to the chapter titled “The PLOT Procedure” in the *Base SAS Procedures Guide* for further information.

**GOUT**=*graphics-catalog*

specifies the graphics catalog for saving traditional graphics output from PROC LIFETEST. The default is WORK.GSEG. The GOUT= option cannot be used if the [LINEPRINTER](#) option or the ODS GRAPHICS ON statement is specified. For more information, refer to the chapter titled “The GREPLAY Procedure” in *SAS/GRAPH Software: Reference*.

**INTERVALS**=*values*

specifies a list of interval endpoints for the life-table method. These endpoints must all be nonnegative numbers. The initial interval is assumed to start at zero whether or not zero is specified in the list. Each interval contains its lower endpoint but does not contain its upper endpoint. When this option is used with METHOD=KM, METHOD=BRESLOW, or METHOD=FH, it reduces the number of survival estimates displayed by showing only the estimates for the smallest time within each specified interval. The INTERVALS= option can be specified in any of the following ways:

list separated by blanks	<b>intervals</b> =1 3 5 7
list separated by commas	<b>intervals</b> =1,3,5,7
<i>x</i> to <i>y</i>	<b>intervals</b> =1 to 7
<i>x</i> to <i>y</i> by <i>z</i>	<b>intervals</b> =1 to 7 by 1
combination of the above	<b>intervals</b> =1,3 to 5,7

For example, the specification

```
intervals=5,10 to 30 by 10
```

produces the set of intervals

$$\{[0, 5), [5, 10), [10, 20), [20, 30), [30, \infty)\}$$

**LANNOTATE=SAS-data-set**

**LANN=SAS-data-set**

specifies an input data set that contains variables for local annotation of traditional graphics. You can use the LANNOTATE= option to specify a different annotation for each BY group, in which case the BY variables must be included in the LANNOTATE= data set. The LANNOTATE= option cannot be used if the **LINEPRINTER** option or the ODS GRAPHICS ON statement is specified. The data set specified must be an **ANNOTATE=** type data set, as described in *SAS/GRAPH Software: Reference*.

If there is no BY-group processing, the **ANNOTATE=** and LANNOTATE= options have the same effects.

**LINEPRINTER**

**LS**

specifies that plots are produced by a line printer instead of by a graphical device.

**MAXTIME=value**

specifies the maximum value of the time variable allowed on the plots so that outlying points do not determine the scale of the time axis of the plots. This option affects only the displayed plots and has no effect on any calculations.

**METHOD=type**

specifies the method to be used to compute the survival function estimates. Valid values for *type* are as follows:

**Breslow**

specifies that the Breslow estimates be computed. The Breslow estimator is the exponentiation of the negative Nelson-Aalen estimator of the cumulative hazard function.

**FH**

specifies that the Fleming-Harrington (FH) estimates be computed. The FH estimator is a tie-breaking modification of the Breslow estimator. If there are no tied event times, this estimator is the same as the Breslow estimator.

**KM**

**PL**

specifies that Kaplan-Meier estimates (also known as the product-limit estimates) be computed.

**ACT**

**LIFE**

**LT**

specifies that life-table estimates (also known as actuarial estimates) be computed.

By default, METHOD=KM.



**MISSING**

allows missing values for numeric variables and blank values for character variables as valid stratum levels. See the section “[Missing Values](#)” on page 3735 for details.

By default, PROC LIFETEST does not use observations with missing values for any stratum variables.

**NELSON****AALEN**

produces the Nelson-Aalen estimates of the cumulative hazards and the corresponding standard errors. This option is ignored if METHOD=LT is specified.

**NINTERVAL=***value*

specifies the number of intervals used to compute the life-table estimates of the survivor function. This parameter is overridden by the [WIDTH=](#) option or the [INTERVALS=](#) option. When you specify the NINTERVAL= option, PROC LIFETEST tries to find an interval that results in round numbers for the endpoints. Consequently, the number of intervals can be different from the number requested. Use the [INTERVALS=](#) option to control the interval endpoints. The default is NINTERVAL=10.

**NOCENSLOT****NOCENS**

requests that the plot of censored observations be suppressed when the [LINEPRINTER](#) and [PLOTS=](#) options are specified. This option is not needed when the life-table method is used to compute the survival estimates, since the plot of censored observations is not produced.

**NOLEFT**

suppress the Number Left column in the survival estimate table. Also suppressed is the Number Event column. This option has no effect for life-table estimate.

**NOPRINT**

suppresses the display of output. This option is useful when only an output data set is needed. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 20, “[Using the Output Delivery System](#)” for more information.

**NOTABLE**

suppresses the display of survival function estimates. Only the number of censored and event times, plots, and test results is displayed.

**OUTSURV=***SAS-data-set***OUTS=***SAS-data-set*

creates an output SAS data set to contain the estimates of the survival function and corresponding confidence limits for all strata. See the section “[OUTSURV= Data Set](#)” on page 3753 for more information about the contents of the OUTSURV= data set.

**OUTTEST=***SAS-data-set***OUTT=***SAS-data-set*

creates an output SAS data set to contain the overall chi-square test statistic for association with failure time for the variables in the TEST statement, the values of the univariate rank test

statistics for each variable in the TEST statement, and the estimated covariance matrix of the univariate rank test statistics. See the section “[OUTTEST= Data Set](#)” on page 3754 for more information about the contents of the OUTTEST= data set.

### ***Line Printer PLOTS= Option***

**PLOTS=***plot-request*

**PLOTS=***(plot-requests)*

controls the line printer plots produced. You must also specify the [LINEPRINTER](#) option to obtain line printer plots. When you specify only one plot-request, you can omit the parentheses around the plot-request. Here are some examples:

```
plots=s
plots=(s ls lls)
```

The *plot-requests* include the following:

#### **CENSORED**

##### **C**

specifies a plot of censored observations. This option is available for METHOD=KM, METHOD=BRESLOW, or METHOD=FM only.

#### **SURVIVAL**

##### **S**

specifies a plot of the estimated SDF versus time.

#### **LOGSURV**

##### **LS**

specifies a plot of the  $-\log(\text{estimated SDF})$  versus time.

#### **LOGLOGS**

##### **LLS**

specifies a plot of the  $\log(-\log(\text{estimated SDF}))$  versus  $\log(\text{time})$ .

#### **HAZARD**

##### **H**

specifies a plot of the estimated hazard function versus time (life-table method only).

#### **PDF**

##### **P**

specifies a plot of the estimated probability density function versus time (life-table method only).

### ***ODS Graphics PLOTS= Option***

**PLOTS**< *(global-plot-option)* > = *plot-request* < *(options)* >

**PLOTS**< *(global-plot-option)* > = *(plot-request* < *(options)* > < ... *plot-request* < *(options)* > >

controls the plots produced using ODS Graphics. When you specify only one *plot-request*, you can omit the parentheses around the *plot-request*. Here are some examples:

```
plots=none
plots=(survival(atrisk=100 to 350 by 50) logsurv)
plots(only)=hazard
```

You must enable ODS Graphics before requesting plots—for example, like this:

```
ods graphics on;

proc lifetest plots=survival(atrisk);
  time T*Status(0);
run;

ods graphics off;
```

For general information about ODS Graphics, see Chapter 21, “[Statistical Graphics Using ODS](#).” If you have enabled ODS Graphics but do not specify the PLOTS= option, then PROC LIFETEST produces a plot of the estimated survivor functions by default. The only global plot option follows:

#### ONLY

specifies that only the specified plots in the list be produced; otherwise, the default survivor function plot is also displayed.

The *plot-requests* and *plot-request* options include the following.

#### ALL

produces all appropriate plots. For METHOD=KM, METHOD=BRESLOW, or METHOD=FH, specifying PLOTS=ALL is equivalent to specifying PLOTS=(SURVIVAL LOGSURV LOGLOGLS HAZARD); for the life-table method, PLOTS=ALL is equivalent to specifying PLOTS=(SURVIVAL LOGSURV LOGLOGS DENSITY HAZARD).

#### HAZARD < (*hazard-options*) >

##### H < *hazard-options* >

plots the estimated hazard functions. Kernel-smoothed estimates are produced for METHOD=KM, METHOD=BRESLOW, or METHOD=FH. You can specify the following *hazard-options*, but only the CL option can be used for the life-table method:

##### **BANDWIDTH**=*bandwidth-option*

##### **BW**=*bandwidth-option*

specifies what bandwidth is chosen for the kernel-smoothing and how it is chosen. You can specify one of the following *bandwidth-options*.

##### *value*

sets the bandwidth to the given *value*.

##### *numeric-list*

selects the bandwidth from the given *numeric-list* that minimizes the mean integrated squared error.

**RANGE**(*lower,upper*)

selects the bandwidth from the interval (*lower, upper*) that minimizes the mean integrated squared error. PROC LIFETEST uses the golden section search algorithm to find the minimum. If there is more than one local minimum in the interval, there is no guarantee that the local minimum found is also the global minimum.

See the section “[Optimal Bandwidth](#)” on page 3746 for details about the mean integrated squared error. If the BANDWIDTH= option is not specified, the default is BANDWIDTH= RANGE(0.2*b*,20*b*), where  $b = \frac{g_u - g_l}{8n^{.2}}$ ,  $g_l$  and  $g_u$  are the values of the GRIDL= and GRIDU= options, respectively, and  $n$  is the total number of noncensored observations.

**GRIDL**=*number*

specifies the lower grid limit for the kernel-smoothed estimate. The default value is the time origin.

**GRIDU**=*number*

specifies the upper grid limit for the kernel-smoothed estimate. The default value equals the maximum event time

**KERNEL**=*kernel-option*

specifies the kernel used. The choices are as follows:

**BIWEIGHT****BW**

$$K_{BW}(x) = \frac{15}{16}(1 - x^2)^2, \quad -1 \leq x \leq 1$$

**EPANECHNIKOV****E**

$$K_E(x) = \frac{3}{4}(1 - x^2), \quad -1 \leq x \leq 1$$

**UNIFORM****U**

$$K_U(x) = \frac{1}{2}, \quad -1 \leq x \leq 1$$

The default is KERNEL=EPANECHNIKOV.

**NMINGRID**=*number*

specifies the number of grid points in determining the mean integrated square error (MISE). The default value is 51.

**NGRID**=*number*

specifies the number of grid points. The default is 101.

**CL**

displays the pointwise confidence limits for the smoothed hazard.

**LOGLOGS****LLS**

plots the log of negative log of estimated survivor functions versus the log of time.

**LOGSURV****LS**

plots the negative log of estimated survivor functions versus time.

**NONE**

suppresses all plots.

**PDF <(CL)>****P <(CL)>**

plots the estimated probability density functions (life-table method only). Pointwise confidence limits are displayed optionally by specifying the CL option.

**SURVIVAL <(survival-options)>****S <(survival-options)>**

plots the estimated survivor functions. Censored times are plotted as a plus sign on the Kaplan-Meier, Breslow, or Fleming-Harrington survival curves unless the NOCENSOR option is specified. You can customize the display by using the following *survival-options*:

**ATRISK <= number-list >**

displays the numbers of subjects at risk at the given times. The *number-list* identifies the times at which the numbers at risk are displayed. If the *number-list* is not specified, PROC LIFETEST uses the default list  $\{0, a, 2a, \dots, n \times a\}$ , where  $a$  and  $n$  are computed by the following algorithm. Let  $m$  be the MAXTIME= value or the largest observed time if the MAXTIME= option is not specified; let  $b = 10^{\text{ceil}[\log_{10}(m)-1]}$ , where  $\text{ceil}()$  is the ceiling function.

$$a = \begin{cases} \frac{b}{2} & \text{if } m < 0.25b \\ 2b & \text{if } m > 0.75b \\ b & \text{otherwise} \end{cases}$$

$$n = \text{integral value of } m/a$$

**CB <=keyword >**

displays the confidence bands (that is, simultaneous confidence intervals) for the survivor functions. You can specify one of the following *keywords*. The default is CB=HW.

**ALL**

displays both the equal-precision and the Hall-Wellner bands.

**EP**

displays the equal-precision band.

**HW**

displays the Hall-Wellner confidence band.

**CL**

displays the pointwise confidence limits for the survivor functions.

**FAILURE****F**

changes all the displays for survivor functions to those for the failure functions. For example, if both the FAILURE and CL options are specified, the plot displays the failure curves as well as the pointwise confidence limits for the failure functions.

**NOCENSOR**

suppresses the plotting of the censored times on a Kaplan-Meier, Breslow, or Fleming-Harrington survival curve.

**STRATA=***strata-option*

specifies how to display the survival/failure curves for multiple strata. This option has no effect if there is only one stratum. You can choose one of the following *strata options*:

**INDIVIDUAL****UNPACK**

specifies that a separate plot be displayed for each stratum.

**OVERLAY**

specifies that the survival/failure curves for the strata be overlaid in one plot.

**PANEL**

specifies that separate plots for the strata be organized into panels of two or four plots, depending on the number of strata.

The default is STRATA=OVERLAY.

**TEST**

displays the *p*-value of a homogeneity test specified in the STRATA statement. If more than one test is produced, the test is chosen in the following order: LOGRANK, WILCOXON, TARONE, PETO, MODPETO, FLEMING, and LR.

*Traditional Graphics PLOTS= Option*

**PLOTS=***plot-request* <(NAME=*name* | '*string*')>

**PLOTS=**(*plot-request* <(NAME=*name* | '*string*')> <, ..., *plot-request* <(NAME=*name* | '*string*')> >)

controls plots produced in traditional graphics. To obtain traditional graphics, you must neither enable ODS Graphics nor specify the **LINEPRINTER** option. For each *plot-request*, you can use the NAME= option to specify a name to identify the plot. The name can be specified as a SAS name or as a quoted string of up to 256 characters. Only the first eight characters are used as the entry name in the **GOUT=** catalog. The *plot-requests* include the following:

**SURVIVAL****S**

plots the estimated survivor functions versus time.

**LOGSURV****LS**

plots the negative log of estimated survivor functions versus time.

**LOGLOGS****LLS**

plots the log of negative log of estimated survivor functions versus the log of time.

**HAZARD****H**

plots estimated hazard function versus time (life-table method only).

**PDF****P**

plots the estimated probability density function versus time (life-table method only).

When you specify only one *plot-request*, you can omit the parentheses around the *plot-request*. Here are some examples:

```
plots=s
plots=(s(name=Surv2), h(name=Haz2))
```

The latter requests a plot of the estimated survivor function versus time and a plot of the estimated hazard function versus time, with Surv2 and Haz2 as their names in the **GOUT=** catalog, respectively.

**REDUCEOUT**

specifies that the **OUTSURV=** data set contain only those observations that are included in the **INTERVALS=** or **TIMELIST=** option. This option has no effect if the **OUTSURV=** option is not specified. It also has no effect if neither the **INTERVALS=** option nor the **TIMELIST=** option is specified.

**SINGULAR=value**

specifies the tolerance for testing singularity of the covariance matrix for the rank test statistics. The test requires that a pivot for sweeping a covariance matrix be at least this number times a norm of the matrix. The default value is 1E-12.

**STDERR**

specifies that the standard error of the survivor function (SDF\_STDERR) be output to the OUTSURV= data set. If the life-table method is used, the standard error of the density function (PDF\_STDERR) and the standard error of the hazard function (HAZ\_STDERR) are also output.

**TIMELIM=***time-limit*

specifies the time limit used in the estimation of the mean survival time and its standard error. The mean survival time can be shown to be the area under the Kaplan-Meier survival curve. However, if the largest observed time in the data is censored, the area under the survival curve is not a closed area. In such a situation, you can choose a time limit  $L$  and estimate the mean survival curve limited to a time  $L$  (Lee 1992, pp. 72–76). This option is ignored if the largest observed time is an event time. Valid *time-limit* values are as follows:

**EVENT****LET**

specifies that the time limit  $L$  be the largest event time in the data. **TIMELIM=EVENT** is the default.

**OBSERVED****LOT**

specifies that the time limit  $L$  be the largest observed time in the data.

*number*

specifies that the time limit  $L$  be the given *number*. The *number* must be positive and at least as large as the largest event time in the data.

**TIMELIST=***number-list*

specifies a list of time points at which the Kaplan-Meier estimates are displayed. The time points are listed in the column labeled *Timelist*. Since the Kaplan-Meier survival curve is a decreasing step function, each given time point falls in an interval that has a constant survival estimate. The event time that corresponds to the beginning of the time interval is displayed along with its survival estimate.

**WIDTH=***value*

sets the width of the intervals used in the life-table calculation of the survival function. This parameter is overridden by the **INTERVALS=** option.

---

## BY Statement

**BY** *variables* ;

You can specify a BY statement with PROC LIFETEST to obtain separate analyses on observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the LIFETEST procedure. The NOTSORTED option does not mean that the data are unsorted but rather that



the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.

- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

The BY statement is more efficient than the STRATA statement for defining strata in large data sets. However, if you use the BY statement to define strata, PROC LIFETEST does not pool over strata for testing the association of survival time with covariates, nor does it test for homogeneity across the BY groups.

When the life-table method is used to estimate survivor functions, each BY group might have a different set of intervals. To make intervals the same across BY groups, use the INTERVALS= or WIDTH= option in the PROC LIFETEST statement.

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

---

## FREQ Statement

**FREQ** *variable* < /option > ;

The FREQ statement identifies a *variable* that contains the frequency of occurrence of each observation. PROC LIFETEST treats each observation as if it appeared  $n$  times, where  $n$  is the value of the FREQ variable for the observation. The FREQ statement is useful for producing life tables when the data are already in the form of a summary data set. If it is not an integer, it is truncated to an integer unless the NOTRUNCATE option is specified. If it is missing or less than or equal zero, the observation is not used.

The following option can be specified in the FREQ statement after a slash (/):

### NOTRUNCATE

#### NOTRUNC

specifies that the frequency values are not truncated to integers. This option does not apply to the Fleming-Harrington estimator (METHOD=FU).

---

## ID Statement

**ID** *variables* ;

The ID statement identifies *variables* whose values are used to label the observations of the Kaplan-Meier, Breslow, or Fleming-Harrington survivor function estimates. SAS format statements can be used to format the values of the ID variables.

## STRATA Statement

**STRATA** *variable* < (*list*) > < ... *variable* < (*list*) > > < /options > ;

The STRATA statement identifies the variables that determine the strata levels. Strata are formed according to the nonmissing values of these variables. The MISSING option can be used to allow missing values as a valid stratum level. Other options enable you to specify various *k*-sample tests, stratified tests, or trend tests and to make multiple-comparison adjustments for paired differences.

In the preceding syntax, *variable* is a variable whose values determine the stratum levels, and *list* is a list of endpoints for a numeric variable. The values for *variable* can be formatted or unformatted. If the variable is a character variable, or if the variable is numeric and no list appears, then the strata are defined by the unique values of the strata variable. More than one variable can be specified in the STRATA statement, and each numeric variable can be followed by a list. Each interval contains its lower endpoint but not its upper endpoint. The corresponding strata are formed by the combination of levels. If a variable is numeric and is followed by a list, then the levels for that variable correspond to the intervals defined by the list. The initial interval is assumed to start at  $-\infty$ , and the final interval is assumed to end at  $\infty$ .

The specification of STRATA variables can have any of the following forms:

list separated by blanks	<b>strata age(5 10 20 30)</b>
list separated by commas	<b>strata age(5,10,20,30)</b>
<i>x</i> to <i>y</i>	<b>strata age(5 to 10)</b>
<i>x</i> to <i>y</i> by <i>z</i>	<b>strata age(5 to 30 by 10)</b>
combination of the above	<b>strata age(5,10 to 50 by 10)</b>

For example, the specification

**strata Age(5,20 to 50 by 10) Sex;**

indicates the following levels for the Age variable:

$\{(-\infty, 5), [5, 20), [20, 30), [30, 40), [40, 50), [50, \infty)\}$

This statement also specifies that the Age strata be further subdivided by values of the variable Sex. In this example, there are 6 age groups by 2 sex groups, forming a total of 12 strata.

The specification of several strata variables, such as

**strata A B C;**

is equivalent to the  $A*B*C$  syntax of the TABLES statement in the FREQ procedure. The number of strata levels usually grows very rapidly with the number of STRATA variables, so you must be cautious when specifying the list of STRATA variables.

When comparing more than two survival curves, a  $k$ -sample test tells you whether the curves are significantly different from each other, but it does not identify which pairs of curves are different. A multiple-comparison adjustment of the  $p$ -values for the paired comparisons retains the same overall false positives as the  $k$ -sample test. Two types of paired comparisons can be made: comparisons between all pairs of curves and comparisons between a control curve and all other curves. You use the **DIFF=** option to specify the comparison type, and you use the **ADJUST=** option to select a method of multiple-comparison adjustments.

Table 49.2 summarizes the options available in the STRATA statement.

**Table 49.2** Options Available in the STRATA Statement

Option	Description
<b>Homogeneity Tests</b>	
<b>GROUP=</b>	Specifies the group variable for stratified tests
<b>NODETAIL</b>	Suppresses printing the test statistic and covariance matrix
<b>NOTEST</b>	Suppresses any tests
<b>TEST=</b>	Specifies tests corresponding to various weight functions
<b>TREND</b>	Requests a trend test
<b>Multiple Comparisons</b>	
<b>ADJUST=</b>	Requests a multiple-comparison adjustment
<b>DIFF=</b>	Specifies the type of differences to consider
<b>Missing Strata Value</b>	
<b>MISSING</b>	Allows missing values as valid stratum values

Options in the STRATA statement are specified after a slash (“/”). The following list describes these options.

#### **ADJUST=method**

specifies the multiple-comparison method for adjusting the  $p$ -values of the paired tests. See the section “[Multiple-Comparison Adjustments](#)” on page 3748 for mathematical details; also see Westfall et al. (1999). The adjustment methods include the following:

#### **BONFERRONI**

##### **BON**

applies the Bonferroni correction to the raw  $p$ -values.

#### **DUNNETT**

performs Dunnett’s two-tailed comparisons of the control group with all other groups. PROC LIFETEST uses the factor-analytic covariance approximation described in Hsu (1992) and identifies the adjustment in the results as “Dunnett-Hsu.” Note that ADJUST=DUNNETT is incompatible with DIFF=ALL.

#### **SCHEFFE**

performs Scheffé’s multiple-comparison adjustment.

**SIDAK**

applies the Šidák correction to the raw  $p$ -values.

**SMM****GTE**

performs the paired comparisons based on the studentized maximum modulus test.

**TUKEY**

performs the paired comparisons based on Tukey's studentized range test. PROC LIFETEST uses the approximation described in Kramer (1956) and identifies the adjustment as "Tukey-Kramer" in the results. Note that ADJUST=TUKEY is incompatible with DIFF=CONTROL.

**SIMULATE** <(simulate-options)>

computes the adjusted  $p$ -values from the simulated distribution of the maximum or maximum absolute value of a multivariate normal random vector. The simulation estimates  $q$ , the true  $(1 - \alpha)$ th quantile, where  $\alpha$  is the value of the ALPHA= *simulate-option*.

The number of samples for the SIMULATE adjustment is set so that the tail area for the simulated  $q$  is within a certain *accuracy radius*  $\gamma$  of  $1 - \alpha$  with an *accuracy confidence* of  $100(1 - \epsilon)\%$ . In equation form,

$$\Pr(|F(\hat{q}) - (1 - \alpha)| \leq \gamma) = 1 - \epsilon$$

where  $\hat{q}$  is the simulated  $q$  and  $F$  is the true distribution function of the maximum; see Edwards and Berry (1987) for details. By default,  $\gamma = 0.005$  and  $\epsilon = 0.01$  so that the tail area of  $\hat{q}$  is within 0.005 of 0.95 with 99% confidence.

The *simulate-options* include the following:

**ACC=value**

specifies the target accuracy radius  $\gamma$  of a  $100(1 - \epsilon)\%$  confidence interval for the true probability content of the estimated  $(1 - \alpha)$ th quantile. The default value is ACC=0.005.

**ALPHA=value**

specifies the value  $\alpha$  for estimating the  $(1 - \alpha)$ th quantile. The default value is the ALPHA= value in the PROC LIFETEST statement, or 0.05 if that option is not specified.

**EPS=value**

specifies the value  $\epsilon$  for a  $100(1 - \epsilon)\%$  confidence interval for the true probability content of the estimated  $(1 - \alpha)$ th quantile. The default value for the accuracy confidence is 99%, corresponding to EPS=0.01.

**NSAMP=n**

specifies the sample size for the simulation. By default,  $n$  is set based on the

values of the target accuracy radius  $\gamma$  and accuracy confidence  $100(1 - \epsilon)\%$  for an interval for the true probability content of the estimated  $(1 - \alpha)$ th quantile. With the default values for  $\gamma$ ,  $\epsilon$ , and  $\alpha$  (0.005, 0.01, and 0.05, respectively), NSAMP=12604 by default.

## REPORT

specifies that a report on the simulation should be displayed, including a listing of the parameters, such as  $\gamma$ ,  $\epsilon$ , and  $\alpha$ , as well as an analysis of various methods for estimating or approximating the quantile.

## SEED=*number*

specifies an integer used to start the pseudo-random number generator for the simulation. If you do not specify a seed, or if you specify a value less than or equal to zero, the seed is by default generated from reading the time of day from the computer's clock.

## DIFF=ALL | CONTROL <('string' <... , 'string')>>

specifies which pairs of survival curves are considered for the multiple comparisons.

### DIFF=ALL

requests all paired comparisons

### DIFF=CONTROL <('string' <... 'string')>>

requests comparisons of the control curve with all other curves. To specify the control curve, you specify the quotes strings of formatted values that represent the curve in parentheses. For example, if Cell='large' identifies the control group, you specify

```
DIFF=CONTROL('large')
```

If more than one variable is used to identify the curves (for example, if Cell='large' and Sex='F' represent the control), you specify

```
DIFF=CONTROL('large' 'F')
```

The order of the quoted strings should correspond to the order of the stratum variables. If no specific curve is specified as the control, the first stratum or group value is used.

By default, DIFF=ALL unless you specify ADJUST=DUNNETT, in which case DIFF=CONTROL.

## GROUP=*variable*

specifies the variable whose formatted values identify the various samples whose underlying survival curves are to be compared. The tests are stratified on the levels of the STRATA variables. For instance, in a multicenter trial in which two forms of therapy are to be compared, you specify the variable identifying therapies as the GROUP= variable and the variable identifying centers as the STRATA variable, in order to perform a stratified test to compare the therapies while controlling the effect of the centers.

**MISSING**

allows missing values to be a stratum level or a valid value of the GROUP= variable.

**NODETAIL**

suppresses the display of the rank statistics and the corresponding covariance matrices for various strata. If the TREND option is specified, the display of the scores for computing the trend tests is suppressed.

**NOTEST**

suppresses the  $k$ -sample tests, stratified tests, and trend tests.

**TREND**

computes the trend tests for testing the null hypothesis that the  $k$  population hazards rate are the same versus an ordered alternatives. If there is only one STRATA variable and the variable is numeric, the unformatted values of the variable are used as the scores; otherwise, the scores are 1, 2, . . . , in the given order of the strata.

**TEST=test-request****TEST=(test-request < . . . test-request > )**

controls the tests produced. Each test corresponds to a different weight function (see the section “[Nonparametric Tests](#)” on page 3747 for the weight functions). The test requests include the following:

ALL	Specifies all the nonparametric tests with $\rho_1=1$ and $\rho_2=0$ for the Fleming and Harrington test—FLEMING(1,0).
FLEMING( $\rho_1, \rho_2$ )	Specifies the family of tests in Harrington and Fleming (1982), where $\rho_1$ and $\rho_2$ are nonnegative numbers. FLEMING( $\rho_1, \rho_2$ ) reduces to the Fleming-Harrington $G^\rho$ family (Fleming and Harrington 1981) when $\rho_2=0$ , which you can specify as FLEMING( $\rho$ ) with one argument. When $\rho=0$ , the test becomes the log-rank test. When $\rho=1$ , the test should be very close to the Peto-Peto test.
LOGRANK	Specifies the log-rank test.
NONE	Suppresses all comparison tests. Specifying TEST=NONE is equivalent to specify NOTEST.
LR	Specifies the likelihood ratio test based on the exponential model.
MODPETO	Specifies the modified Peto-Peto test.
PETO	Specifies the Peto-Peto test. The test is also referred to as the Peto-Peto-Prentice test.
WILCOXON	Specifies the Wilcoxon test. The test is also referred to as the Gehan test or the Breslow test.
TARONE	Specifies the Tarone-Ware test.

By default, TEST=(LOGRANK WILCOXON LR) for the  $k$ -sample tests, and TEST=(LOGRANK WILCOXON) for stratified and trend tests.

---

## TEST Statement

**TEST** *variables* ;

The TEST statement specifies a list of numeric covariates (prognostic variables) that you want tested for association with the failure time.

Two sets of rank statistics are computed. These rank statistics and their variances are pooled over all strata. Univariate (marginal) test statistics are displayed for each of the covariates.

Additionally, a sequence of test statistics for joint effects of covariates is displayed. The first element of the sequence is the largest univariate test statistic. Other variables are then added on the basis of the largest increase in the joint test statistic. The process continues until all the variables have been added or until the remaining variables are linearly dependent on the previously added variables.

See the section “[Rank Tests for the Association of Survival Time with Covariates](#)” on page 3750 for more information.

---

## TIME Statement

**TIME** *variable* < \**censor(list)* > ;

The TIME statement is required. It is used to indicate the failure time variable, where *variable* is the name of the failure time variable that can be optionally followed by an asterisk, the name of the censoring variable, and a parenthetical list of values that correspond to right censoring. The censoring values should be numeric, nonmissing values. For example, the statement

```
time T*Flag(1,2);
```

identifies the variable T as containing the observed failure times (event or censored). If the variable Flag has the value 1 or 2, the corresponding value of T is a right-censored value.

---

## Details: LIFETEST Procedure

---

### Missing Values

Observations with a missing value for either the failure time or the censoring variable are not used in the analysis. If a stratum variable value is missing, the observation is not used; however, the MISSING option can be used to request that missing values be treated as valid stratum values. If any variable specified in the TEST statement has a missing value, that observation is not used in the calculation of the rank statistics.

## Computational Formulas

### Breslow, Fleming-Harrington, and Kaplan-Meier Methods

Let  $t_1 < t_2 < \dots < t_D$  represent the distinct event times. For each  $i = 1, \dots, D$ , let  $n_i$  be the number of surviving units (the size of the risk set) just prior to  $t_i$ . Let  $d_i$  be the number of units that fail at  $t_i$ , and let  $s_i = n_i - d_i$ . If the NOTRUNCATE option is specified in the FREQ statement,  $n_i$ ,  $d_i$ , and  $s_i$  can be nonintegers.

The Breslow estimate of the survivor function is

$$\hat{S}(t_i) = \exp\left(-\sum_{j=1}^i \frac{d_j}{n_j}\right)$$

Note that the Breslow estimate is the exponentiation of the negative Nelson-Aalen estimate of the cumulative hazard function.

The Fleming-Harrington estimate (Fleming and Harrington 1984) of the survivor function is

$$\hat{S}(t_i) = \exp\left(-\sum_{k=1}^i \sum_{j=0}^{d_k-1} \frac{1}{n_k - j}\right)$$

If the frequency values are not integers, the Fleming-Harrington estimate cannot be computed.

The Kaplan-Meier (product-limit) estimate of the survivor function at  $t_i$  is the cumulative product

$$\hat{S}(t_i) = \prod_{j=1}^i \left(1 - \frac{d_j}{n_j}\right)$$

Notice that all the estimators are defined to be right continuous; that is, the events at  $t_i$  are included in the estimate of  $S(t_i)$ . The corresponding estimate of the standard error is computed using Greenwood's formula (Kalbfleisch and Prentice 1980) as

$$\hat{\sigma}(\hat{S}(t_i)) = \hat{S}(t_i) \sqrt{\sum_{j=1}^i \frac{d_j}{n_j s_j}}$$

The first quartile (or the 25th percentile) of the survival time is the time beyond which 75% of the subjects in the population under study are expected to survive. It is estimated by

$$q_{.25} = \min\{t_j | \hat{S}(t_j) < 0.75\}$$

If  $\hat{S}(t)$  is exactly equal to 0.75 from  $t_j$  to  $t_{j+1}$ , the first quartile is taken to be  $(t_j + t_{j+1})/2$ . If it happens that  $\hat{S}(t)$  is greater than 0.75 for all values of  $t$ , the first quartile cannot be estimated and is represented by a missing value in the printed output.

The general formula for estimating the 100 $p$ th percentile point is

$$q_p = \min\{t_j | \hat{S}(t_j) < 1 - p\}$$



The second quartile (the median) and the third quartile of survival times correspond to  $p=0.5$  and  $p=0.75$ , respectively.

Brookmeyer and Crowley (1982) have constructed the confidence interval for the median survival time based on the confidence interval for the  $S(t)$ . The methodology is generalized to construct the confidence interval for the 100 $p$ th percentile based on a  $g$ -transformed confidence interval for  $S(t)$  (Klein and Moeschberger 1997). You can use the CONFTYPE= option to specify the  $g$ -transformation. The 100(1 -  $\alpha$ )% confidence interval for the first quartile survival time is the set of all points  $t$  that satisfy

$$\left| \frac{g(\hat{S}(t)) - g(1 - 0.25)}{g'(\hat{S}(t))\hat{\sigma}(\hat{S}(t))} \right| \leq z_{1-\frac{\alpha}{2}}$$

where  $g'(x)$  is the first derivative of  $g(x)$  and  $z_{1-\frac{\alpha}{2}}$  is the 100(1 -  $\frac{\alpha}{2}$ )th percentile of the standard normal distribution.

Consider the bone marrow transplant data described in [Example 49.2](#). The following table illustrates the construction of the confidence limits for the first quartile in the ALL group. Values of  $\frac{g(\hat{S}(t)) - g(1-0.25)}{g'(\hat{S}(t))\hat{\sigma}(\hat{S}(t))}$  that lie between  $\pm z_{1-\frac{0.05}{2}} = \pm 1.965$  are highlighted.

Constructing 95% Confidence Limits for the 25th Percentile							
t	$\hat{S}(t)$	$\hat{\sigma}(\hat{S}(t))$	$\frac{g(\hat{S}(t)) - g(1-0.25)}{g'(\hat{S}(t))\hat{\sigma}(\hat{S}(t))}$				
			LINEAR	LOGLOG	LOG	ASINSQRT	LOGIT
1	0.97368	0.025967	8.6141	2.37831	9.7871	4.44648	2.47903
55	0.94737	0.036224	5.4486	2.36375	6.1098	3.60151	2.46635
74	0.92105	0.043744	3.9103	2.16833	4.3257	2.94398	2.25757
86	0.89474	0.049784	2.9073	<b>1.89961</b>	3.1713	2.38164	1.97023
104	0.86842	0.054836	2.1595	<b>1.59196</b>	2.3217	<b>1.87884</b>	<b>1.64297</b>
107	0.84211	0.059153	<b>1.5571</b>	<b>1.26050</b>	<b>1.6490</b>	<b>1.41733</b>	<b>1.29331</b>
109	0.81579	0.062886	<b>1.0462</b>	<b>0.91307</b>	<b>1.0908</b>	<b>0.98624</b>	<b>0.93069</b>
110	0.78947	0.066135	<b>0.5969</b>	<b>0.55415</b>	<b>0.6123</b>	<b>0.57846</b>	<b>0.56079</b>
122	0.73684	0.071434	<b>-0.1842</b>	<b>-0.18808</b>	<b>-0.1826</b>	<b>-0.18573</b>	<b>-0.18728</b>
129	0.71053	0.073570	<b>-0.5365</b>	<b>-0.56842</b>	<b>-0.5222</b>	<b>-0.54859</b>	<b>-0.56101</b>
172	0.68421	0.075405	<b>-0.8725</b>	<b>-0.95372</b>	<b>-0.8330</b>	<b>-0.90178</b>	<b>-0.93247</b>
192	0.65789	0.076960	<b>-1.1968</b>	<b>-1.34341</b>	<b>-1.1201</b>	<b>-1.24712</b>	<b>-1.30048</b>
194	0.63158	0.078252	<b>-1.5133</b>	<b>-1.73709</b>	<b>-1.3870</b>	<b>-1.58613</b>	<b>-1.66406</b>
230	0.60412	0.079522	<b>-1.8345</b>	-2.14672	<b>-1.6432</b>	<b>-1.92995</b>	-2.03291
276	0.57666	0.080509	-2.1531	-2.55898	<b>-1.8825</b>	-2.26871	-2.39408
332	0.54920	0.081223	-2.4722	-2.97389	-2.1070	-2.60380	-2.74691
383	0.52174	0.081672	-2.7948	-3.39146	-2.3183	-2.93646	-3.09068
418	0.49428	0.081860	-3.1239	-3.81166	-2.5177	-3.26782	-3.42460
466	0.46682	0.081788	-3.4624	-4.23445	-2.7062	-3.59898	-3.74781
487	0.43936	0.081457	-3.8136	-4.65971	-2.8844	-3.93103	-4.05931
526	0.41190	0.080862	-4.1812	-5.08726	-3.0527	-4.26507	-4.35795
609	0.38248	0.080260	-4.5791	-5.52446	-3.2091	-4.60719	-4.64271
662	0.35306	0.079296	-5.0059	-5.96222	-3.3546	-4.95358	-4.90900

Consider the LINEAR transformation where  $g(x) = x$ . The event times that satisfy

$\left| \frac{g(\hat{S}(t)) - g(1-p)}{g'(\hat{S}(t)) \sqrt{\hat{V}(\hat{S}(t))}} \right| \leq 1.9599$  include 107, 109, 110, 122, 129, 172, 192, 194, and 230. The confidence of the interval [107, 230] is less than 95%. Brookmeyer and Crowley (1982) suggest extending the confidence interval to but not including the next event time. As such the 95% confidence interval for the first quartile based on the linear transform is [107, 276). The following table lists the confidence intervals for the various transforms.

95% CI's for the 25th Percentile		
CONFTYPE	[Lower	Upper)
LINEAR	107	276
LOGLOG	86	230
LOG	107	332
ASINSQRT	104	276
LOGIT	104	230

Sometimes, the confidence limits for the quartiles cannot be estimated. For convenience of explanation, consider the linear transform  $g(x) = x$ . If the curve representing the upper confidence limits for the survivor function lies above 0.75, the upper confidence limit for first quartile cannot be estimated. On the other hand, if the curve representing the lower confidence limits for the survivor function lies above 0.75, the lower confidence limit for the quartile cannot be estimated.

The estimated mean survival time is

$$\hat{\mu} = \sum_{i=1}^D \hat{S}(t_{i-1})(t_i - t_{i-1})$$

where  $t_0$  is defined to be zero. When the largest observed time is censored, this sum underestimates the mean. The standard error of  $\hat{\mu}$  is estimated as

$$\hat{\sigma}(\hat{\mu}) = \sqrt{\frac{m}{m-1} \sum_{i=1}^{D-1} \frac{A_i^2}{n_i s_i}}$$

where

$$A_i = \sum_{j=i}^{D-1} \hat{S}(t_j)(t_{j+1} - t_j)$$

$$m = \sum_{j=1}^D d_j$$

If the largest observed time is not an event, you can use the TIMELIM= option to specify a time limit  $L$  and estimate the mean survival time limited to the time  $L$  and its standard error by replacing  $k$  by  $k + 1$  with  $t_{k+1} = L$ .

### Nelson-Aalen Estimate of the Cumulative Hazard Function

The Nelson-Aalen cumulative hazard estimator, defined up to the largest observed time on study, is

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

and its estimated variance is

$$\hat{V}(\tilde{H}(t)) = \sum_{t_i \leq t} \frac{d_i}{n_i^2}$$

### Life-Table Method

The life-table estimates are computed by counting the numbers of censored and uncensored observations that fall into each of the time intervals  $[t_{i-1}, t_i)$ ,  $i = 1, 2, \dots, k + 1$ , where  $t_0 = 0$  and  $t_{k+1} = \infty$ . Let  $n_i$  be the number of units entering the interval  $[t_{i-1}, t_i)$ , and let  $d_i$  be the number of events occurring in the interval. Let  $b_i = t_i - t_{i-1}$ , and let  $n'_i = n_i - w_i/2$ , where  $w_i$  is the number of units censored in the interval. The *effective sample size* of the interval  $[t_{i-1}, t_i)$  is denoted by  $n'_i$ . Let  $t_{mi}$  denote the midpoint of  $[t_{i-1}, t_i)$ .

The conditional probability of an event in  $[t_{i-1}, t_i)$  is estimated by

$$\hat{q}_i = \frac{d_i}{n'_i}$$

and its estimated standard error is

$$\hat{\sigma}(\hat{q}_i) = \sqrt{\frac{\hat{q}_i \hat{p}_i}{n'_i}}$$

where  $\hat{p}_i = 1 - \hat{q}_i$ .

The estimate of the survival function at  $t_i$  is

$$\hat{S}(t_i) = \begin{cases} 1 & i = 0 \\ \hat{S}(t_{i-1}) \hat{p}_i & i > 0 \end{cases}$$

and its estimated standard error is

$$\hat{\sigma}(\hat{S}(t_i)) = \hat{S}(t_i) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n'_j \hat{p}_j}}$$

The density function at  $t_{mi}$  is estimated by

$$\hat{f}(t_{mi}) = \frac{\hat{S}(t_i) \hat{q}_i}{b_i}$$

and its estimated standard error is

$$\hat{\sigma}(\hat{f}(t_{mi})) = \hat{f}(t_{mi}) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n'_j \hat{p}_j} + \frac{\hat{p}_i}{n'_i \hat{q}_i}}$$

The estimated hazard function at  $t_{mi}$  is

$$\hat{h}(t_{mi}) = \frac{2\hat{q}_i}{b_i(1 + \hat{p}_i)}$$

and its estimated standard error is

$$\hat{\sigma}(\hat{h}(t_{mi})) = \hat{h}(t_{mi}) \sqrt{\frac{1 - (b_i \hat{h}(t_{mi})/2)^2}{n'_i \hat{q}_i}}$$

Let  $[t_{j-1}, t_j]$  be the interval in which  $\hat{S}(t_{j-1}) \geq \hat{S}(t_i)/2 > \hat{S}(t_j)$ . The median residual lifetime at  $t_i$  is estimated by

$$\hat{M}_i = t_{j-1} - t_i + b_j \frac{\hat{S}(t_{j-1}) - \hat{S}(t_i)/2}{\hat{S}(t_{j-1}) - \hat{S}(t_j)}$$

and the corresponding standard error is estimated by

$$\hat{\sigma}(\hat{M}_i) = \frac{\hat{S}(t_i)}{2\hat{f}(t_{mj})\sqrt{n'_i}}$$

### Interval Determination

If you want to determine the intervals exactly, use the INTERVALS= option in the PROC LIFETEST statement to specify the interval endpoints. Use the WIDTH= option to specify the width of the intervals, thus indirectly determining the number of intervals. If neither the INTERVALS= option nor the WIDTH= option is specified in the life-table estimation, the number of intervals is determined by the NINTERVAL= option. The width of the time intervals is 2, 5, or 10 times an integer (possibly a negative integer) power of 10. Let  $c = \log_{10}(\text{maximum observed time/number of intervals})$ , and let  $b$  be the largest integer not exceeding  $c$ . Let  $d = 10^{c-b}$  and let

$$a = 2 \times I(d \leq 2) + 5 \times I(2 < d \leq 5) + 10 \times I(d > 5)$$

with  $I$  being the indicator function. The width is then given by

$$\text{width} = a \times 10^b$$

By default, NINTERVAL=10.

### Pointwise Confidence Limits in the OUTSURV= Data Set

Pointwise confidence limits are computed for the survivor function, and for the density function and hazard function when the life-table method is used. Let  $\alpha$  be specified by the ALPHA= option. Let  $z_{\alpha/2}$  be the critical value for the standard normal distribution. That is,  $\Phi(-z_{\alpha/2}) = \alpha/2$ , where  $\Phi$  is the cumulative distribution function of the standard normal random variable.

### Survivor Function

When the computation of confidence limits for the survivor function  $S(t)$  is based on the asymptotic normality of the survival estimator  $\hat{S}(t)$ , the approximate confidence interval might include impossible values outside the range  $[0,1]$  at extreme values of  $t$ . This problem can be avoided by applying the asymptotic normality to a transformation of  $S(t)$  for which the range is unrestricted. In addition, certain transformed confidence intervals for  $S(t)$  perform better than the usual linear confidence intervals (Borgan and Liestøl 1990). The CONFTYPE= option enables you to pick one of the following transformations: the log-log function (Kalbfleisch and Prentice 1980), the arcsine-square root function (Nair 1984), the logit function (Meeker and Escobar 1998), the log function, and the linear function.

Let  $g$  be the transformation that is being applied to the survivor function  $S(t)$ . By the delta method, the standard error of  $g(\hat{S}(t))$  is estimated by

$$\tau(t) = \hat{\sigma} \left[ g(\hat{S}(t)) \right] = g'(\hat{S}(t)) \hat{\sigma}[\hat{S}(t)]$$

where  $g'$  is the first derivative of the function  $g$ . The  $100(1-\alpha)\%$  confidence interval for  $S(t)$  is given by

$$g^{-1} \left\{ g[\hat{S}(t)] \pm z_{\frac{\alpha}{2}} g'[\hat{S}(t)] \hat{\sigma}[\hat{S}(t)] \right\}$$

where  $g^{-1}$  is the inverse function of  $g$ .

**Arcsine-Square Root Transformation** The estimated variance of  $\sin^{-1}(\sqrt{\hat{S}(t)})$  is  $\hat{\tau}^2(t) = \frac{\hat{\sigma}^2[\hat{S}(t)]}{4\hat{S}(t)[1-\hat{S}(t)]}$ . The  $100(1-\alpha)\%$  confidence interval for  $S(t)$  is given by

$$\sin^2 \left\{ \max \left[ 0, \sin^{-1}(\sqrt{\hat{S}(t)}) - z_{\frac{\alpha}{2}} \hat{\tau}(t) \right] \right\} \leq S(t) \leq \sin^2 \left\{ \min \left[ \frac{\pi}{2}, \sin^{-1}(\sqrt{\hat{S}(t)}) + z_{\frac{\alpha}{2}} \hat{\tau}(t) \right] \right\}$$

**Linear Transformation** This is the same as having no transformation in which  $g$  is the identity. The  $100(1-\alpha)\%$  confidence interval for  $S(t)$  is given by

$$\hat{S}(t) - z_{\frac{\alpha}{2}} \hat{\sigma}[\hat{S}(t)] \leq S(t) \leq \hat{S}(t) + z_{\frac{\alpha}{2}} \hat{\sigma}[\hat{S}(t)]$$

**Log Transformation** The estimated variance of  $\log(\hat{S}(t))$  is  $\hat{\tau}^2(t) = \frac{\hat{\sigma}^2(\hat{S}(t))}{\hat{S}^2(t)}$ . The  $100(1-\alpha)\%$  confidence interval for  $S(t)$  is given by

$$\hat{S}(t) \exp \left( -z_{\frac{\alpha}{2}} \hat{\tau}(t) \right) \leq S(t) \leq \hat{S}(t) \exp \left( z_{\frac{\alpha}{2}} \hat{\tau}(t) \right)$$

**Log-Log Transformation** The estimated variance of  $\log(-\log(\hat{S}(t)))$  is  $\hat{\tau}^2(t) = \frac{\hat{\sigma}^2[\hat{S}(t)]}{[\hat{S}(t) \log(\hat{S}(t))]^2}$ . The  $100(1-\alpha)\%$  confidence interval for  $S(t)$  is given by

$$\left[ \hat{S}(t) \right]^{\exp \left( z_{\frac{\alpha}{2}} \hat{\tau}(t) \right)} \leq S(t) \leq \left[ \hat{S}(t) \right]^{\exp \left( -z_{\frac{\alpha}{2}} \hat{\tau}(t) \right)}$$

**Logit Transformation** The estimated variance of  $\log\left(\frac{\hat{S}(t)}{1-\hat{S}(t)}\right)$  is  $\hat{\tau}^2(t) = \frac{\hat{\sigma}^2(\hat{S}(t))}{\hat{S}^2(t)[1-\hat{S}(t)]^2}$ . The 100(1- $\alpha$ )% confidence limits for  $S(t)$  are given by

$$\frac{\hat{S}(t)}{\hat{S}(t) + [1 - \hat{S}(t)] \exp\left(z_{\frac{\alpha}{2}} \hat{\tau}(t)\right)} \leq S(t) \leq \frac{\hat{S}(t)}{\hat{S}(t) + [1 - \hat{S}(t)] \exp\left(-z_{\frac{\alpha}{2}} \hat{\tau}(t)\right)}$$

### Density and Hazard Functions

For the life-table method, a 100(1- $\alpha$ )% confidence interval for hazard function or density function at time  $t$  is computed as

$$\hat{g}(t) \pm z_{\alpha/2} \hat{\sigma}[\hat{g}(t)]$$

where  $\hat{g}(t)$  is the estimate of either the hazard function or the density function at time  $t$ , and  $\hat{\sigma}[\hat{g}(t)]$  is the corresponding standard error estimate.

### Simultaneous Confidence Intervals for Kaplan-Meier Curves

The pointwise confidence interval for the survivor function  $S(t)$  is valid for a single fixed time at which the inference is to be made. In some applications, it is of interest to find the upper and lower confidence bands that guarantee, with a given confidence level, that the survivor function falls within the band for all  $t$  in some interval. Hall and Wellner (1980) and Nair (1984) provide two different approaches for deriving the confidence bands. An excellent review can be found in Klein and Moeschberger (1997). You can use the CONFBAND= option in the SURVIVAL statement to select the confidence bands. The EP confidence band provides confidence bounds that are proportional to the pointwise confidence interval, while those of the HW band are not proportional to the pointwise confidence bounds. The maximum time,  $t_U$ , for the bands can be specified by the BANDMAX= option; the minimum time,  $t_L$ , can be specified by the BANDMIN= option. Transformations that are used to improve the pointwise confidence intervals can be applied to improve the confidence bands. It might turn out that the upper and lower bounds of the confidence bands are not decreasing in  $t_L < t < t_U$ , which is contrary to the nonincreasing characteristic of survivor function. Meeker and Escobar (1998) suggest making an adjustment so that the bounds do not increase: if the upper bound is increasing on the right, it is made flat from the minimum to  $t_U$ ; if the lower bound is increasing from the right, it is made flat from  $t_L$  to the maximum. PROC LIFETEST does not make any adjustment for the nondecreasing behavior of the confidence bands in the OUTSURV= data set. However, the adjustment was made in the display of the confidence bands by using ODS Graphics.

For Kaplan-Meier estimation, let  $t_1 < t_2 < \dots < t_D$  be the  $D$  distinct events times, and at time  $t_i$ , there are  $d_i$  events. Let  $Y_i$  be the number of individuals who are at risk at time  $t_i$ . The variance of  $\hat{S}(t)$ , given by the Greenwood formula, is  $\hat{\sigma}^2[\hat{S}(t)] = \sigma_S^2(t) \hat{S}^2(t)$ , where

$$\sigma_S^2(t) = \sum_{t_i \leq t} \frac{d_i}{Y_i(Y_i - d_i)}$$

Let  $t_L < t_U$  be the time range for the confidence band so that  $t_U$  is less than or equal to the largest event time. For the Hall-Wellner band,  $t_L$  can be zero, but for the equal-precision band,  $t_L$  is greater

than or equal to the smallest event time. Let

$$a_L = \frac{n\sigma_S^2(t_L)}{1 + n\sigma_S^2(t_L)} \text{ and } a_U = \frac{n\sigma_S^2(t_U)}{1 + n\sigma_S^2(t_U)}$$

Let  $\{W^0(u), 0 \leq u \leq 1\}$  be a Brownian bridge.

### **Hall-Wellner Band**

The  $100(1-\alpha)\%$  HW band of Hall and Wellner (1980) is

$$\hat{S}(t) - h_\alpha(a_L, a_U)n^{-\frac{1}{2}}[1 + n\sigma_S^2(t)]\hat{S}(t) \leq S(t) \leq \hat{S}(t) + h_\alpha(a_L, a_U)n^{-\frac{1}{2}}[1 + n\sigma_S^2(t)]\hat{S}(t)$$

for all  $t_L \leq t \leq t_U$ , where  $h_\alpha(a_L, a_U)$  is given by

$$\alpha = \Pr\left\{\sup_{a_L \leq u \leq a_U} |W^0(u)| > h_\alpha(a_L, a_U)\right\}$$

The critical values are computed from the results in Chung (1986).

Note that the given confidence band has a formula similar to that of the (linear) pointwise confidence interval, where  $h_\alpha(a_L, a_U)$  and  $n^{-\frac{1}{2}}[1 + n\sigma_S^2(t)]\hat{S}(t)$  in the former correspond to  $z_{\frac{\alpha}{2}}$  and  $\hat{\sigma}(\hat{S}(t))$  in the latter, respectively. You can obtain the other transformations (arcsine-square root, log-log, log, and logit) for the confidence bands by replacing  $z_{\frac{\alpha}{2}}$  and  $\hat{\tau}(t)$  in the corresponding pointwise confidence interval formula by  $h_\alpha(a_L, a_U)$  and the following  $\hat{\tau}(t)$ , respectively.

Arcsine-Square Root Transformation

$$\hat{\tau}(t) = \frac{1 + n\sigma_S^2(t)}{2} \sqrt{\frac{S(t)}{n[1 - S(t)]}}$$

Log Transformation

$$\hat{\tau}(t) = \frac{1 + n\sigma_S^2(t)}{\sqrt{n}}$$

Log-Log Transformation

$$\hat{\tau}(t) = \frac{1 + n\sigma_S^2(t)}{\sqrt{n} |\log[\hat{S}(t)]|}$$

Logit Transformation

$$\hat{\tau}(t) = \frac{1 + n\sigma_S^2(t)}{\sqrt{n}[1 - \hat{S}(t)]}$$

### Equal-Precision Band

The 100(1- $\alpha$ )% EP band of Nair (1984) is

$$\hat{S}(t) - e_\alpha(a_L, a_U) \hat{S}(t) \sigma_S(t) \leq S(t) \leq \hat{S}(t) + e_\alpha(a_L, a_U) \hat{S}(t) \sigma_S(t)$$

for all  $t_L \leq t \leq t_U$ , where  $e_\alpha(a_L, a_U)$  is given by

$$\alpha = \Pr\left\{ \sup_{a_L \leq u \leq a_U} \frac{|W^0(u)|}{[u(1-u)]^{\frac{1}{2}}} > e_\alpha(a_L, a_U) \right\}$$

PROC LIFETEST uses the approximation of Miller and Siegmund (1982, Equation 8) to approximate the tail probability in which  $e_\alpha(a_L, a_U)$  is obtained by solving  $x$  in

$$\frac{4x\phi(x)}{x} + \phi(x) \left( x - \frac{1}{x} \right) \log \left[ \frac{a_U(1-a_L)}{a_L(1-a_U)} \right] = \alpha$$

where  $\phi(x)$  is the standard normal density function evaluated at  $x$ . Note that the confidence bounds given are proportional to the pointwise confidence intervals. As a matter of fact, this confidence band and the (linear) pointwise confidence interval have the same formula except for the critical values ( $z_{\frac{\alpha}{2}}$  for the pointwise confidence interval and  $e_\alpha(a_L, a_U)$  for the band). You can obtain the other transformations (arcsine-square root, log-log, log, and logit) for the confidence bands by replacing  $z_{\frac{\alpha}{2}}$  by  $e_\alpha(a_L, a_U)$  in the formula of the pointwise confidence intervals.

### Kernel-Smoothed Hazard Estimate

Kernel-smoothed estimators of the hazard function  $h(t)$  are based on the Nelson-Aalen estimator  $\tilde{H}(t)$  and its variance  $\hat{V}(\tilde{H}(t))$ . Consider the jumps of  $\tilde{H}(t)$  and  $\hat{V}(\tilde{H}(t))$  at the event times  $t_1 < t_2 < \dots < t_D$  as follows:

$$\begin{aligned} \Delta \tilde{H}(t_i) &= \tilde{H}(t_i) - \tilde{H}(t_{i-1}) \\ \hat{V}(\tilde{H}(t_i)) &= \hat{V}(\tilde{H}(t_i)) - \hat{V}(\tilde{H}(t_{i-1})) \end{aligned}$$

where  $t_0=0$ .

The kernel-smoothed estimator of  $h(t)$  is a weighted average of  $\Delta \tilde{H}(t)$  over event times that are within a bandwidth distance  $b$  of  $t$ . The weights are controlled by the choice of kernel function,  $K()$ , defined on the interval  $[-1,1]$ . The choices of  $K()$  are as follows:

*uniform kernel*

$$K_U(x) = \frac{1}{2}, \quad -1 \leq x \leq 1$$

*Epanechnikov kernel*

$$K_E(x) = \frac{3}{4}(1-x^2), \quad -1 \leq x \leq 1$$



*biweight* kernel

$$K_{BW}(x) = \frac{15}{16}(1-x^2)^2, \quad -1 \leq x \leq 1$$

The kernel-smoothed hazard rate estimator is defined for all time points on  $(0, t_D)$ . For time points  $t$  for which  $b \leq t \leq t_D - b$ , the kernel-smoothed estimated of  $h(t)$  based on the kernel  $K()$  is given by

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^D K\left(\frac{t-t_i}{b}\right) \Delta \tilde{H}(t_i)$$

The variance of  $\hat{h}(t)$  is estimated by

$$\hat{\sigma}^2(\hat{h}(t)) = \frac{1}{b^2} \sum_{i=1}^D K\left(\frac{t-t_i}{b}\right)^2 \Delta \hat{V}(\tilde{H}(t_i))$$

For  $t < b$ , the symmetric kernels  $K()$  are replaced by the corresponding asymmetric kernels of Gasser and Müller (1979). Let  $q = \frac{t}{b}$ . The modified kernels are as follows:

*uniform* kernel

$$K_{U,q}(x) = \frac{4(1+q^3)}{(1+q)^4} + \frac{6(1-q)}{(1+q)^3}x, \quad -1 \leq x \leq q$$

*Epanechnikov* kernel

$$K_{E,q}(x) = K_E(x) \frac{64(2-4q+6q^2-3q^3) + 240(1-q)^2x}{(1+q)^4(19-18q+3q^2)}, \quad -1 \leq x \leq q$$

*biweight* kernel

$$K_{BW}(x) = K_{BW} \frac{64(8-24q+48q^2-45q^3+15q^4) + 1120(1-q)^3x}{(1+q)^5(81-168q+126q^2-40q^3+5q^4)}, \quad -1 \leq x \leq q$$

For  $t_D - b \leq t \leq t_D$ , let  $q = \frac{t_D-t}{b}$ . The asymmetric kernels for  $t < b$  are used with  $x$  replaced by  $-x$ .

Using the log transform on the smoothed hazard rate, the  $100(1-\alpha)\%$  pointwise confidence interval for the smoothed hazard rate  $h(t)$  is given by

$$\hat{h}(t) = \hat{h}(t) \exp \left[ \pm \frac{z_{1-\alpha/2} \hat{\sigma}(\hat{h}(t))}{\hat{h}(t)} \right]$$

where  $z_{1-\frac{\alpha}{2}}$  is the  $100(1-\frac{\alpha}{2})$ th percentile of the standard normal distribution.

### Optimal Bandwidth

The following mean integrated squared error (MISE) over the range  $\tau_L$  and  $\tau_U$  is used as a measure of the global performance of the kernel function estimator

$$\begin{aligned} MISE(b) &= E \int_{\tau_L}^{\tau_U} (\hat{h}(u) - h(u))^2 du \\ &= E \int_{\tau_L}^{\tau_U} \hat{h}^2(u) du - 2E \int_{\tau_L}^{\tau_U} \hat{h}(u)h(u) du + E \int_{\tau_L}^{\tau_U} h^2(u) du \end{aligned}$$

The last term is independent of the choice of the kernel and bandwidth and can be ignored when you are looking for the best value of  $b$ . The first integral can be approximated by using the trapezoid rule by evaluating  $\hat{h}(t)$  at a grid of points  $\tau_L = u_1 < \dots < u_M = \tau_U$ . You can specify  $\tau_L$ ,  $\tau_R$ , and  $M$  by using the options MISEMIN=, MISEMAX=, and MISENUM=, respectively, of the HAZARD plot. The second integral can be estimated by Ramlau-Hansen (1983a, b) cross-validation estimate

$$\frac{1}{b} \sum_{i \neq j} K\left(\frac{t_i - t_j}{b}\right) \Delta \hat{H}(t_i) \Delta \hat{H}(t_j)$$

Therefore, for a fixed kernel, the optimal bandwidth is the quantity  $b$  that minimizes

$$g(b) = \sum_{i=1}^{M-1} \left[ \frac{u_{i+1} - u_i}{2} \left( \hat{h}^2(u_i) + \hat{h}^2(u_{i+1}) \right) \right] - \frac{2}{b} \sum_{i \neq j} K\left(\frac{t_i - t_j}{b}\right) \Delta \hat{H}(t_i) \Delta \hat{H}(t_j)$$

The minimization is carried out by the golden section search algorithm.

### Comparison of Two or More Groups of Survival Data

Let  $k$  be the number of groups. Let  $S_i(t)$  be the underlying survivor function  $i$ th group,  $i = 1, \dots, k$ . The null and alternative hypotheses to be tested are

$$H_0 : S_1(t) = S_2(t) = \dots = S_k(t) \text{ for all } t \leq \tau$$

versus

$$H_1 : \text{at least one of the } S_i(t) \text{'s is different for some } t \leq \tau$$

respectively, where  $\tau$  is the largest observed time.

### Likelihood Ratio Test

The likelihood ratio test statistic (Lawless 1982) for test  $H_0$  versus  $H_1$  assumes that the data in the various samples are exponentially distributed and tests that the scale parameters are equal. The test statistic is computed as

$$\chi^2 = 2N \log \left( \frac{T}{N} \right) - 2 \sum_{j=1}^k N_j \log \left( \frac{T_j}{N_j} \right)$$

where  $N_j$  is the total number of events in the  $j$ th stratum,  $N = \sum_{j=1}^k N_j$ ,  $T_j$  is the total time on test in the  $j$ th stratum, and  $T = \sum_{j=1}^k T_j$ . The approximate probability value is computed by treating  $\chi^2$  as having a chi-square distribution with  $k-1$  degrees of freedom.

### Nonparametric Tests

Let  $t_1 < t_2 < \dots < t_D$  be the distinct event times in the pooled sample. At time  $t_i$ , let  $W(t_i)$  be a positive weight function, and let  $n_{ij}$  and  $d_{ij}$  be the size of the risk set and the number of events in the  $j$ th sample, respectively. Let  $n_i = \sum_{j=1}^k n_{ij}$ ,  $d_i = \sum_{j=1}^k d_{ij}$ , and  $s_i = n_i - d_i$ .

The choices of the weight function  $W(t_i)$  are given in Table 49.3.

**Table 49.3** Weight Functions for Various Tests

Test	$W(t_i)$
Log-rank	1.0
Wilcoxon	$n_i$
Tarone-Ware	$\sqrt{n_i}$
Peto-Peto	$\tilde{S}(t_i)$
Modified Peto-Peto	$\tilde{S}(t_i) \frac{n_i}{n_i + 1}$
Harrington-Fleming ( $p, q$ )	$[\hat{S}(t_i)]^p [1 - \hat{S}(t_i)]^q, p \geq 0, q \geq 0$

where  $\hat{S}(t)$  is the product-limit estimate at  $t$  for the pooled sample, and  $\tilde{S}(t)$  is a survivor function estimate close to  $\hat{S}(t)$  given by

$$\tilde{S}(t) = \prod_{t_i \leq t} \left( 1 - \frac{d_i}{n_i + 1} \right)$$

**Unstratified Tests** The rank statistics (Klein and Moeschberger 1997, Section 7.3) for testing  $H_0$  versus  $H_1$  have the form of a  $k$ -vector  $\mathbf{v} = (v_1, v_2, \dots, v_k)'$  with

$$v_j = \sum_{i=1}^D W(t_i) \left\{ d_{ij} - \frac{n_{ij} d_i}{n_i} \right\}$$

and the estimated covariance matrix,  $\mathbf{V} = (V_{jl})$ , is given by

$$V_{jl} = \sum_{i=1}^D W^2(t_i) \left\{ \frac{d_i s_i (n_i n_{il} \delta_{jl} - n_{ij} n_{il})}{n_i^2 (n_i - 1)} \right\}$$

where  $\delta_{jl}$  is 1 if  $j = l$  and 0 otherwise. The term  $v_j$  can be interpreted as a weighted sum of observed minus expected numbers of failure under the null hypothesis of identical survival curves. The overall test statistic for homogeneity is  $\mathbf{v}' \mathbf{V}^- \mathbf{v}$ , where  $\mathbf{V}^-$  denotes a generalized inverse of  $\mathbf{V}$ . This statistic is treated as having a chi-square distribution with degrees of freedom equal to the rank of  $\mathbf{V}$  for the purposes of computing an approximate probability level.

**Stratified Tests** Suppose the test is to be stratified on  $M$  levels of a set of STRATA variables. Based only on the data of the  $s$ th stratum ( $s = 1 \dots M$ ), let  $\mathbf{v}_s$  be the test statistic (Klein and Moeschberger 1997, Section 7.5) for the  $s$ th stratum, and let  $\mathbf{V}_s$  be its covariance matrix. Let

$$\begin{aligned}\mathbf{v} &= \sum_{s=1}^M \mathbf{v}_s \\ \mathbf{V} &= \sum_{s=1}^M \mathbf{V}_s\end{aligned}$$

A global test statistic is constructed as

$$\chi^2 = \mathbf{v}'\mathbf{V}\mathbf{v}$$

Under the null hypothesis, the test statistic has a  $\chi^2$  distribution with the same degrees of freedom as the individual test for each stratum.

**Multiple-Comparison Adjustments** Let  $\chi_r^2$  denote a chi-squared random variable with  $r$  degrees of freedom. Denote  $\phi$  and  $\Phi$  as the density function and the cumulative distribution function of a standard normal distribution, respectively. Let  $m$  be the number of comparisons; that is,

$$m = \begin{cases} \frac{k(k-1)}{2} & \text{DIFF = ALL} \\ k-1 & \text{DIFF = CONTROL} \end{cases}$$

For a two-sided test comparing the survival of the  $j$ th group with that of  $l$ th group,  $1 \leq j \neq l \leq r$ , the test statistic is

$$z_{jl}^2 = \frac{(v_j - v_l)^2}{V_{jj} + V_{ll} - 2V_{jl}}$$

and the raw  $p$ -value is

$$p = \Pr(\chi_1^2 > z_{jl}^2)$$

Adjusted  $p$ -values for various multiple-comparison adjustments are computed as follows:

*Bonferroni Adjustment*

$$p = \min\{1, m\Pr(\chi_1^2 > z_{jl}^2)\}$$

*Dunnett-Hsu Adjustment*

With the first group being the control, let  $\mathbf{C} = (c_{ij})$  be the  $(r-1) \times r$  matrix of contrasts; that is,

$$c_{ij} = \begin{cases} 1 & i = 1, \dots, r-1, j = 2, \dots, r \\ -1 & j = i+1, i = 2, \dots, r \\ 0 & \text{otherwise} \end{cases}$$

Let  $\mathbf{\Sigma} \equiv (\sigma_{ij})$  and  $\mathbf{R} \equiv (r_{ij})$  be covariance and correlation matrices of  $\mathbf{Cv}$ , respectively; that is,

$$\mathbf{\Sigma} = \mathbf{CVC}'$$

and

$$r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

The factor-analytic covariance approximation of Hsu (1992) is to find  $\lambda_1, \dots, \lambda_{r-1}$  such that

$$\mathbf{R} = \mathbf{D} + \mathbf{\lambda}\mathbf{\lambda}'$$

where  $\mathbf{D}$  is a diagonal matrix with the  $j$ th diagonal element being  $1 - \lambda_j$  and  $\mathbf{\lambda} = (\lambda_1, \dots, \lambda_{r-1})'$ . The adjusted  $p$ -value is

$$p = 1 - \int_{-\infty}^{\infty} \phi(y) \prod_{i=1}^{r-1} \left[ \Phi\left(\frac{\lambda_i y + z_{jl}}{\sqrt{1 - \lambda_i^2}}\right) - \Phi\left(\frac{\lambda_i y - z_{jl}}{\sqrt{1 - \lambda_i^2}}\right) \right] dy$$

which can be obtained in a DATA step as

$$p = \text{PROBMC}(\text{"DUNNETT2"}, z_{ij}, \dots, r-1, \lambda_1, \dots, \lambda_{r-1}).$$

*Scheffé Adjustment*

$$p = \Pr(\chi_{r-1}^2 > z_{jl}^2)$$

*Šidák Adjustment*

$$p = 1 - \{1 - \Pr(\chi_1^2 > z_{jl}^2)\}^m$$

*SMM Adjustment*

$$p = 1 - [2\Phi(z_{jl}) - 1]^m$$

which can also be evaluated in a DATA step as

$$p = 1 - \text{PROBMC}(\text{"MAXMOD"}, z_{jl}, \dots, m).$$

*Tukey Adjustment*

$$p = 1 - \int_{-\infty}^{\infty} r \phi(y) [\Phi(y) - \Phi(y - \sqrt{2}z_{jl})]^{r-1} dy$$

which can also be evaluated in a DATA step as

$$p = 1 - \text{PROBMC}(\text{"RANGE"}, \sqrt{2}z_{jl}, \dots, r).$$

**Trend Tests** Trend tests (Klein and Moeschberger 1997, Section 7.4) have more power to detect ordered alternatives as

$$H_2 : S_1(t) \geq S_2(t) \geq \dots \geq S_k(t), t \leq \tau, \text{ with at least one inequality}$$

or

$$H_2 : S_1(t) \leq S_2(t) \leq \dots \leq S_k(t), t \leq \tau, \text{ with at least one inequality}$$

Let  $a_1 < a_2 < \dots < a_k$  be a sequence of scores associated with the  $k$  samples. The test statistic and its standard error are given by  $\sum_{j=1}^k a_j v_j$  and  $\sum_{j=1}^k \sum_{l=1}^k a_j a_l V_{jl}$ , respectively. Under  $H_0$ , the  $z$ -score

$$Z = \frac{\sum_{j=1}^k a_j v_j}{\sqrt{\sum_{j=1}^k \sum_{l=1}^k a_j a_l V_{jl}}}$$

has, asymptotically, a standard normal distribution. PROC LIFETEST provides both one-tail and two-tail  $p$ -values for the test.

## Rank Tests for the Association of Survival Time with Covariates

The rank tests for the association of covariates (Kalbfleisch and Prentice 1980, Chapter 6) are more general cases of the rank tests for homogeneity. In this section, the index  $\alpha$  is used to label all observations,  $\alpha = 1, 2, \dots, n$ , and the indices  $i, j$  range only over the observations that correspond to events,  $i, j = 1, 2, \dots, k$ . The ordered event times are denoted as  $t_{(j)}$ , the corresponding vectors of covariates are denoted as  $\mathbf{z}_{(i)}$ , and the ordered times, both censored and event times, are denoted as  $t_\alpha$ .

The rank test statistics have the form

$$\mathbf{v} = \sum_{\alpha=1}^n c_{\alpha, \delta_\alpha} \mathbf{z}_\alpha$$

where  $n$  is the total number of observations,  $c_{\alpha, \delta_\alpha}$  are rank scores, which can be either log-rank or Wilcoxon rank scores,  $\delta_\alpha$  is 1 if the observation is an event and 0 if the observation is censored, and  $\mathbf{z}_\alpha$  is the vector of covariates in the TEST statement for the  $\alpha$ th observation. Notice that the scores,  $c_{\alpha, \delta_\alpha}$ , depend on the censoring pattern and that the terms are summed up over all observations.

The log-rank scores are

$$c_{\alpha, \delta_\alpha} = \sum_{(j: t_{(j)} \leq t_\alpha)} \left( \frac{1}{n_j} - \delta_\alpha \right)$$

and the Wilcoxon scores are

$$c_{\alpha, \delta_\alpha} = 1 - (1 + \delta_\alpha) \prod_{(j: t_{(j)} \leq t_\alpha)} \frac{n_j}{n_j + 1}$$

where  $n_j$  is the number at risk just prior to  $t_{(j)}$ .

The estimates used for the covariance matrix of the log-rank statistics are

$$\mathbf{V} = \sum_{i=1}^k \frac{\mathbf{V}_i}{n_i}$$

where  $\mathbf{V}_i$  is the corrected sum of squares and crossproducts matrix for the risk set at time  $t_{(i)}$ ; that is,

$$\mathbf{V}_i = \sum_{(\alpha: t_\alpha \geq t_{(i)})} (\mathbf{z}_\alpha - \bar{\mathbf{z}}_i)' (\mathbf{z}_\alpha - \bar{\mathbf{z}}_i)$$

where

$$\bar{\mathbf{z}}_i = \sum_{(\alpha: t_\alpha \geq t_{(i)})} \frac{\mathbf{z}_\alpha}{n_i}$$

The estimate used for the covariance matrix of the Wilcoxon statistics is

$$\mathbf{V} = \sum_{i=1}^k \left[ a_i (1 - a_i^*) (2\mathbf{z}_{(i)} \mathbf{z}_{(i)}' + \mathbf{S}_i) - (a_i^* - a_i) \left( a_i \mathbf{x}_i \mathbf{x}_i' + \sum_{j=i+1}^k a_j (\mathbf{x}_i \mathbf{x}_j' + \mathbf{x}_j \mathbf{x}_i') \right) \right]$$

where

$$a_i = \prod_{j=1}^i \frac{n_j}{n_j + 1}$$

$$a_i^* = \prod_{j=1}^i \frac{n_j + 1}{n_j + 2}$$

$$\mathbf{S}_i = \sum_{(\alpha: t_{(i+1)} > t_\alpha > t_{(i)})} \mathbf{z}_\alpha \mathbf{z}_\alpha'$$

$$\mathbf{x}_i = 2\mathbf{z}_{(i)} + \sum_{(\alpha: t_{(i+1)} > t_\alpha > t_{(i)})} \mathbf{z}_\alpha$$

In the case of tied failure times, the statistics  $\mathbf{v}$  are averaged over the possible orderings of the tied failure times. The covariance matrices are also averaged over the tied failure times. Averaging the covariance matrices over the tied orderings produces functions with appropriate symmetries for the tied observations; however, the actual variances of the  $\mathbf{v}$  statistics would be smaller than the preceding estimates. Unless the proportion of ties is large, it is unlikely that this will be a problem.

The univariate tests for each covariate are formed from each component of  $\mathbf{v}$  and the corresponding diagonal element of  $\mathbf{V}$  as  $v_i^2 / V_{ii}$ . These statistics are treated as coming from a chi-square distribution for calculation of probability values.

The statistic  $\mathbf{v}'\mathbf{V}^{-1}\mathbf{v}$  is computed by sweeping each pivot of the  $\mathbf{V}$  matrix in the order of greatest increase to the statistic. The corresponding sequence of partial statistics is tabulated. Sequential

increments for including a given covariate and the corresponding probabilities are also included in the same table. These probabilities are calculated as the tail probabilities of a chi-square distribution with one degree of freedom. Because of the selection process, these probabilities should not be interpreted as *p*-values.

If desired for data screening purposes, the output data set requested by the OUTTEST= option can be treated as a sum of squares and crossproducts matrix and processed by the REG procedure by using the option METHOD=RSQUARE. Then the sets of variables of a given size can be found that give the largest test statistics. [Output 49.1](#) illustrates this process.

---

## Computer Resources

The data are first read and sorted into strata. If the data are originally sorted by failure time and censoring state, with smaller failure times coming first and event values preceding censored values in cases of ties, the data can be processed by strata without additional sorting. Otherwise, the data are read into memory by strata and sorted.

## Memory Requirements

For a given BY group, define the following:

$N$	the total number of observations
$V$	the number of STRATA variables
$C$	the number of covariates listed in the TEST statement
$L$	total length of the ID variables in bytes
$S$	number of strata
$n$	maximum number of observations within strata

$b$	$12 + 8C + L$
$m1$	$(112 + 16V) \times S$
$m2$	$50 \times b \times S$
$m3$	$(50 + n) \times (b + 4)$
$m4$	$8(C + 4)^2$
$m5$	$20N + 8S \times (S + 4)$



The memory, in bytes, required to process the BY group is at least

$$m1 + \max(m2, m3) + m4$$

The test of equality of survival functions across strata requires additional memory ( $m5$  bytes). However, if this additional memory is not available, PROC LIFETEST skips the test for equality of survival functions and finishes the other computations. Additional memory is required for the PLOTS= option. Temporary storage of  $16n$  bytes is required to store the product-limit estimates for plotting.

---

## Output Data Sets

### OUTSURV= Data Set

You can specify the OUTSURV= option in the PROC LIFETEST statement to create an output data set containing the following columns:

- any specified BY variables
- any specified STRATA variables, their values coming from either their original values or the midpoints of the stratum intervals if endpoints are used to define strata (semi-infinite intervals are labeled by their finite endpoint)
- STRATUM, a numeric variable that numbers the strata
- the time variable as given in the TIME statement. For METHOD=KM, METHOD=BRESLOW, or METHOD=FH, it contains the observed failure or censored times. For the life-table estimates, it contains the lower endpoints of the time intervals.
- SURVIVAL, a variable containing the survivor function estimates
- CONFTYPE, a variable containing the name of the transformation applied to the survival time in the computation of confidence intervals (if the OUT= option is specified in the SURVIVAL statement)
- SDF\_LCL, a variable containing the lower limits of the pointwise confidence intervals for the survivor function
- SDF\_UCL, a variable containing the upper limits of the pointwise confidence intervals for the survivor function

If the estimation uses the product-limit, Breslow, or Fleming-Harrington method, then the data set also contains the following:

- \_CENSOR\_, an indicator variable that has a value 1 for a censored observation and a value 0 for an event observation

- SDF\_STDERR, a variable containing the standard error of the survivor function estimator (if the STDERR option is specified in the PROC LIFETEST statement)
- HW\_LCL, a variable containing the lower limits of the Hall-Wellner confidence bands (if the CONFBAND=HW option or the CONFBAND=ALL option is specified in the PROC LIFETEST statement)
- HW\_UCL, a variable containing the upper limits of the Hall-Wellner confidence bands (if the CONFBAND=HW option or the CONFBAND=ALL option is specified in the PROC LIFETEST statement)
- EP\_LCL, a variable containing the lower limits of the equal-precision confidence bands (if the CONFBAND=EP option or the CONFBAND=ALL option is specified in the PROC LIFETEST statement)
- EP\_UCL, a variable containing the upper limits of the equal-precision confidence bands (if the CONFBAND=EP option or the CONFBAND=ALL option is specified in the PROC LIFETEST statement)

If the estimation uses the life-table method, then the data set also contains the following:

- MIDPOINT, a variable containing the value of the midpoint of the time interval
- PDF, a variable containing the density function estimates
- PDF\_LCL, a variable containing the lower endpoints of the PDF confidence intervals
- PDF\_UCL, a variable containing the upper endpoints of the PDF confidence intervals
- HAZARD, a variable containing the hazard estimates
- HAZ\_LCL, a variable containing the lower endpoints of the hazard confidence intervals
- HAZ\_UCL, a variable containing the upper endpoints of the hazard confidence intervals

Each survival function contains an initial observation with the value 1 for the SDF and the value 0 for the time. The output data set contains an observation for each distinct failure time if the product-limit, Breslow, or Fleming-Harrington method is used, or it contains an observation for each time interval if the life-table method is used. The product-limit, Breslow, or Fleming-Harrington survival estimates are defined to be right continuous; that is, the estimates at a given time include the factor for the failure events that occur at that time.

Labels are assigned to all the variables in the output data set except the BY variable and the STRATA variable.

## OUTTEST= Data Set

The OUTTEST= option in the LIFETEST statement creates an output data set containing the rank statistics for testing the association of failure time with covariates. It contains the following:

- any specified BY variables
- `_TYPE_`, a character variable of length 8 that labels the type of rank test, either “LOG-RANK” or “WILCOXON”
- `_NAME_`, a character variable of length 8 that labels the rows of the covariance matrix and the test statistics
- the `TIME` variable, containing the overall test statistic in the observation that has `_NAME_` equal to the name of the time variable and the univariate test statistics under their respective covariates.
- all variables listed in the `TEST` statement

The output is in the form of a symmetric matrix formed by the covariance matrix of the rank statistics bordered by the rank statistics and the overall chi-square statistic. If the value of `_NAME_` is the name of a variable in the `TEST` statement, the observation contains a row of the covariance matrix and the value of the rank statistic in the time variable. If the value of `_NAME_` is the name of the `TIME` variable, the observation contains the values of the rank statistics in the variables from the `TEST` list and the value of the overall chi-square test statistic in the `TIME` variable.

Two complete sets of statistics labeled by the `_TYPE_` variable are produced, one for the log-rank test and one for the Wilcoxon test.

---

## Displayed Output

If you use the `NOPRINT` option in the `PROC LIFETEST` statement, the procedure does not display any output.

## Product-Limit Survival Estimates

The “Product-Limit Survival Estimates” table is displayed if you request the product-limit method of estimation. The table displays the following:

- the observed (event or censored) time
- the number of units at risk (if you specify the `ATRISK` option in the `PROC LIFETEST` statement)
- the number of events (if you specify the `ATRISK` option in the `PROC LIFETEST` statement)
- the product-limit estimate of the survivor function
- the corresponding estimate of the cumulative distribution function of the failure time
- the standard error estimate of the survivor function estimator

- the Nelson-Aalen cumulative hazard function estimate (if the NELSON option is specified in the PROC LIFETEST statement)
- the standard error of the Nelson-Aalen estimator (if the NELSON option is specified in the PROC LIFETEST statement)
- the number of event times that have been observed
- the number of event or censored times that remain to be observed
- the frequency of the observed times (if you specify the FREQ statement)
- values of the ID variables (if you specify the ID statement)

For ODS purposes, the name of this table is “ProductLimitEstimates.”

### Breslow Survival Estimates

The “Breslow Survival Estimates” table is displayed if you request the Breslow method of estimation. The table displays the following:

- the observed (event or censored) time
- the number of units at risk (if you specify the ATRISK option in the PROC LIFETEST statement)
- the number of events (if you specify the ATRISK option in the PROC LIFETEST statement)
- the Breslow estimate of the survivor function
- the corresponding estimate of the cumulative distribution function of the failure time
- the standard error estimate of the survivor function estimator
- the Nelson-Aalen cumulative hazard function estimate (if the NELSON option is specified in the PROC LIFETEST statement)
- the standard error of the Nelson-Aalen estimator (if the NELSON option is specified in the PROC LIFETEST statement)
- the number of event times that have been observed
- the number of event or censored times that remain to be observed
- the frequency of the observed times (if you specify the FREQ statement)
- values of the ID variables (if you specify the ID statement)

For ODS purposes, the name of this table is “BreslowEstimates.”

## Fleming-Harrington Survival Estimates

The “Fleming-Harrington Survival Estimates” table is displayed if you request the Fleming-Harrington method of estimation. The table displays the following:

- the observed (event or censored) time
- the number of units at risk (if you specify the ATRISK option in the PROC LIFETEST statement)
- the number of events (if you specify the ATRISK option in the PROC LIFETEST statement)
- the Fleming-Harrington estimate of the survivor function
- the corresponding estimate of the cumulative distribution function of the failure time
- the standard error estimate of the survivor function estimator
- the Nelson-Aalen cumulative hazard function estimate (if the NELSON option is specified in the PROC LIFETEST statement)
- the standard error of the Nelson-Aalen estimator (if the NELSON option is specified in the PROC LIFETEST statement)
- the number of event times that have been observed
- the number of event or censored times that remain to be observed
- the frequency of the observed times (if you specify the FREQ statement)
- values of the ID variables (if you specify the ID statement)

For ODS purposes, the name of this table is “FlemingEstimates.”

## Quartile Estimates

The “Quartiles Estimates” table is displayed if you request the product-limit, Breslow, or Fleming-Harrington method of estimation. The table displays the following:

- point estimates of the quartiles of the survival times
- the lower and upper confidence limits for the quartiles

For ODS purposes, the name of this table is “Quartiles.”

## Mean Estimate

The “Mean Estimate” table is displayed if you request the product-limit, Breslow, or Fleming-Harrington method of estimation. The table displays the following:

- the estimated mean survival time
- the estimated standard error of the mean estimator

For ODS purposes, the name of this table is “Means.”

## Life-Table Survival Estimates

The “Life-Table Survival Estimates” table is displayed if you request the life-table method of estimation. The table displays the following:

- the time intervals into which the failure and censored times are distributed. Each interval is from the lower limit, up to but not including the upper limit; if the upper limit is infinity, the missing value is printed.
- the number of events that occur in the interval
- the number of censored observations that fall into the interval
- the effective sample size for the interval
- the estimate of conditional probability of events (failures) in the interval
- the standard error of the conditional probability estimator
- the estimate of the survival function at the beginning of the interval
- the estimate of the cumulative distribution function of the failure time at the beginning of the interval
- the standard error estimate of the survivor function estimator
- the estimate of the median residual lifetime, which is the amount of time elapsed before reducing the number of at-risk units to one-half. This is also known as the *median future lifetime* in Elandt-Johnson and Johnson (1980)).
- the estimated standard error of the median residual lifetime estimator
- the density function estimated at the midpoint of the interval
- the standard error estimate of the density estimator
- the hazard rate estimated at the midpoint of the interval
- the standard error estimate of the hazard estimator

For ODS purposes, the name of this table is “LifetableEstimates.”

## Summary of the Number of Censored and Uncensored Values

The “Summary of the Number of Censored and Uncensored Values” table displays following:

- the stratum identification (if the STRATA statement is specified)
- the total number of observations,
- the number of event observations,
- the number of censored observations, and
- the percentage of censored observations.

For ODS purposes, the name of this table is “CensoredSummary.”

## Rank Statistics

The “Rank Statistics” table contains the test statistics of the nonparametric  $k$ -sample tests. For ODS purposes, the name of this table is “HomStats.”

## Covariance Matrix for the Log-Rank Statistics

The “Covariance Matrix for the Log-Rank Statistics” table is displayed if the log-rank  $k$ -sample test is requested. For ODS purposes, the name of this table is “LogrankHomCov.”

## Covariance Matrix for the Wilcoxon Statistics

The “Covariance Matrix for the Wilcoxon Statistics” table is displayed if the Wilcoxon  $k$ -sample test is requested. For ODS purposes, the name of this table is “WilHomCov.”

## Covariance Matrix for the Tarone Statistics

The “Covariance Matrix for the Tarone Statistics” table is displayed if the Tarone-Ware  $k$ -sample test is requested. For ODS purposes, the name of this table is “TaroneHomCov.”

## Covariance Matrix for the Peto Statistics

The “Covariance Matrix for the Peto Statistics” table is displayed if the Peto-Peto  $k$ -sample test is requested. For ODS purposes, the name of this table is “PetoHomCov.”

### Covariance Matrix for the ModPeto Statistics

The “Covariance Matrix for the ModPeto Statistics” table is displayed if the modified Peto-Peto  $k$ -sample test is requested. For ODS purposes, the name of this table is “ModPetoHomCov.”

### Covariance Matrix for the Fleming Statistics

The “Covariance Matrix for the Fleming Statistics” table is displayed if the Fleming-Harrington  $k$ -sample test is requested. For ODS purposes, the name of this table is “FlemingHomCov.”

### Test of Equality over Strata

The “Test of Equality over Strata” table is displayed if an unstratified  $k$ -sample test is carried out. The table contains the chi-square statistics, degrees of freedom, and  $p$ -values of the nonparametric tests and the likelihood ratio test (which is based on the exponential distribution). For ODS purposes, the name of this table is “HomTests.”

### Stratified Test of Equality over Group

The “Stratified Test of Equality over Group” table is displayed if a stratified test is carried out. The tables contains the chi-square statistics, degrees of freedom, and  $p$ -values of the stratified tests. For ODS purposes, the name of this table is “HomTests.”

### Scores for Trend Test

The “Scores for Trend Test” table is displayed if the TREND option is specified in the STRATA statement. The table contains the set of scores used to construct the trend tests. For ODS purposes, the name of this table is “TrendScores.”

### Trend Tests

The “Trend Tests” table is displayed if the TREND option is specified in the STRATA statement. The table contains the results of the trend tests. For ODS purposes, the name of this table is “TrendTests.”

### Adjustment for Multiple Comparisons for the Log-Rank Test

The “Adjustment for Multiple Comparisons for the Log-Rank Test” table is displayed if the log-rank test and a multiple-comparison adjustment method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”



### Adjustment for Multiple Comparisons for the Wilcoxon Test

The “Adjustment for Multiple Comparisons for the Wilcoxon Test” table is displayed if the Wilcoxon test and a multiple-comparison method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”

### Adjustment for Multiple Comparisons for the Tarone Test

The “Adjustment for Multiple Comparisons for the Tarone Test” table is displayed if the Tarone-Ware test and a multiple-comparison method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”

### Adjustment for Multiple Comparisons for the Peto Test

The “Adjustment for Multiple Comparisons for the Peto Test” table is displayed if the Peto-Peto test and a multiple-comparison method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”

### Adjustment for Multiple Comparisons for the ModPeto Test

The “Adjustment for Multiple Comparisons for the ModPeto Test” table is displayed if the modified Peto-Peto test and a multiple-comparison method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”

### Adjustment for Multiple Comparisons for the Fleming Test

The “Adjustment for Multiple Comparisons for the Fleming Test” table is displayed if the Fleming-Harrington test and a multiple-comparison method are specified. The table contains the chi-square statistics and the raw and adjusted  $p$ -values of the paired comparisons. For ODS purposes, the name of this table is “SurvDiff.”

### Univariate Chi-Squares for the Log-Rank Test

The “Univariate Chi-Squares for the Log-Rank Test” table is displayed if the TEST statement is specified. The table displays the log-rank test results for individual variables in the TEST statement. For ODS purposes, the name of this table is “LogUniChiSq.”

### Covariance Matrix of the Log-Rank Statistics

The “Covariance Matrix of the Log-Rank Statistics” table is displayed if the TEST statement is specified. The table displays the estimated covariance matrix of the log-rank statistics for association. For ODS purposes, the name of this table is “LogTestCov.”

### Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test

The “Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test” table is displayed if the TEST statement is specified. The table contains the sequence of partial chi-square statistics for the log-rank test in the order of the greatest increase to the overall test statistic, the degrees of freedom of the partial chi-square statistics, the approximate probability values of the partial chi-square statistics, the chi-square increments for including the given variables, and the probability values of the chi-square increments. For ODS purposes, the name of this table is “LogForStepSeq.”

### Univariate Chi-Squares for the Wilcoxon Test

The “Univariate Chi-Squares for the Wilcoxon Test” table displays the Wilcoxon test results for individual variables in the TEST statement. For ODS purposes, the name of this table is “WilUniChiSq.”

### Covariance Matrix of the Wilcoxon Statistics

The “Covariance Matrix of the Wilcoxon Statistics” table is displayed if the TEST statement is specified. The table displays the estimated covariance matrix of the Wilcoxon statistics for association. For ODS purposes, the name of this table is “WilTestCov.”

### Forward Stepwise Sequence of Chi-Squares for the Wilcoxon Test

The “Forward Stepwise Sequence of Chi-Squares for the Wilcoxon Test” table is displayed if the TEST statement is specified. The table contains the sequence of partial chi-square statistics for the Wilcoxon test in the order of the greatest increase to the overall test statistic, the degrees of freedom of the partial chi-square statistics, the approximate probability values of the partial chi-square statistics, the chi-square increments for including the given variables, and the probability values of the chi-square increments. For ODS purposes, the name of this table is “WilForStepSeq.”

---

## ODS Table Names

PROC LIFETEST assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in [Table 49.4](#). For more information about ODS, see Chapter 20, “[Using the Output Delivery System](#).”

**Table 49.4** ODS Tables Produced by PROC LIFETEST

ODS Table Name	Description	Statement / Option
BreslowEstimates	Breslow estimates	PROC LIFETEST METHOD=B
CensoredSummary	Number of event and censored observations	PROC LIFETEST METHOD=PL   B   FH
FlemingEstimates	Fleming-Harrington estimates	PROC LIFETEST METHOD=FH
FlemingHomCov	Covariance matrix for $k$ -sample FLEMING statistics	STRATA / TEST=FLEMING
HomStats	Test statistics for $k$ -sample tests	STRATA / TEST=
HomTests	Results of $k$ -sample tests	STRATA / TEST=
LifetableEstimates	Life-table survival estimates	PROC LIFETEST METHOD=LT
LogForStepSeq	Forward stepwise sequence for the log-rank statistics for association	TEST /
LogHomCov	Covariance matrix for $k$ -sample LOGRANK statistics	STRATA / TEST=LOGRANK
LogTestCov	Covariance matrix for log-rank statistics for association	TEST /
LogUniChisq	Univariate chi-squares for log-rank statistics for association	TEST /
Means	Mean and standard error of survival times	PROC LIFETEST METHOD=PL
ModPetoHomCov	Covariance matrix for $k$ -sample MODPETO statistics	STRATA / TEST=MODPETO
PetoHomCov	Covariance matrix for $k$ -sample PETO statistics	STRATA / TEST=PETO
ProductLimitEstimates	Product-limit survival estimates	PROC LIFETEST METHOD=PL
Quartiles	Quartiles of the survival times	PROC LIFETEST METHOD=PL   B   FH
SurvDiff	Adjustments for multiple comparisons	STRATA / ADJUST= & DIFF=
TaroneHomCov	Covariance matrix for $k$ -sample TARONE statistics	STRATA / TEST=TARONE
TrendScores	Scores used to construct trend tests	STRATA / TREND
TrendTests	Results of trend tests	STRATA / TREND
WilForStepSeq	Forward stepwise sequence for the log-rank statistics for association	TEST /
WilHomCov	Covariance matrix for $k$ -sample WILCOXON statistics	STRATA / TEST=WILCOXON
WilTestCov	Covariance matrix for log-rank statistics for association	TEST /

**Table 49.4** *continued*

ODS Table Name	Description	Statement / Option
WilUniChiSq	Univariate chi-squares for TEST / Wilcoxon statistics for association	

## ODS Graphics

This section describes the use of ODS for creating statistical graphs with the LIFETEST procedure. To request these graphs you must specify the ODS GRAPHICS ON statement. See Chapter 21, “Statistical Graphics Using ODS,” for more information. The survival plot is produced by default; other graphs are produced by using the PLOTS= option in the PROC LIFETEST statement. You can reference every graph produced through ODS Graphics with a name. The names of the graphs that PROC LIFETEST generates are listed in Table 49.5, along with the required keywords for the PLOTS= option.

**Table 49.5** ODS Graphics Produced by PROC LIFETEST

ODS Graph Name	Plot Description	PLOTS=Option
DensityPlot	Estimated density for life-table method	PDF
FailurePlot	Estimated failure function	SURVIVAL(FAILURE)
HazardPlot	Estimated hazard function for life-table method or smoothed hazard for product-limit, Breslow, or Fleming-Harrington method	HAZARD
LogNegLogSurvivalPlot	Log of negative log of the estimated survivor function	LOGLOGS
NegLogSurvivalPlot	Negative log of the estimated survivor function	LOGSURV
SurvivalPlot	Estimated survivor function	SURVIVAL
SurvivalPlot	Estimated survivor function with number of subjects at risk	SURVIVAL(ATRISK=)
SurvivalPlot	Estimated survivor function with point-wise confidence limits	SURVIVAL(CL)
SurvivalPlot	Estimated survivor function with equal-precision band	SURVIVAL(CB=EP)
SurvivalPlot	Estimated survivor function with Hall-Wellner band	SURVIVAL(CB=HW)
SurvivalPlot	Estimated survivor function with homogeneity test $p$ -value	SURVIVAL(TEST)

## Additional Dynamic Variables for Survival Plots Using ODS Graphics

PROC LIFETEST passes a number of summary statistics as dynamic variables to the ODS Graphics for survival plots. Table 49.6 and Table 49.7 list these additional dynamic variables for the Kaplan-Meier curves and the life-table curves, respectively. These dynamic variables are not declared in the templates for the survival curves, but you can declare them and use them to enhance the default plots. The names of the dynamic variables depend on the STRATA= suboption of the PLOTS=SURVIVAL option: STRATA=INDIVIDUAL produces a separate plot for each stratum, and STRATA=OVERALL produces one plot with overlaid curves.

**Table 49.6** Additional Dynamic Variables for `Stat.Graphics.ProductLimitSurvival`

STRATA=	Dynamic	Description
OVERLAY	StrVal $j$	Label for the $j$ th stratum
	NObs $j$	Number of observations in the $j$ th stratum
	NEvent $j$	Number of events in the $j$ th stratum
	Median $j$	Median survival time of the $j$ th stratum
	LowerMedian $j$	Lower median survival time of the $j$ th stratum
	UpperMedian $j$	Upper median survival time of the $j$ th stratum
	pctMedianConfid	Percent confidence of the median intervals
INDIVIDUAL	NObs	Number of observations
	NEvent	Number of events
	Median	Median survival time
	LowerMedian	Lower median survival time
	UpperMedian	Upper median survival time
	pctMedianConfid	Percent confidence of the median interval

**Table 49.7** Additional Dynamic Variables for `Stat.Graphics.LifetableSurvival`

STRATA=	Dynamic	Description
OVERLAY	StrVal $j$	Label for the $j$ th stratum
	NObs $j$	Number of observations in the $j$ th stratum
	NEvent $j$	Number of events in the $j$ th stratum
INDIVIDUAL	NObs	Number of observations
	NEvent	Number of events

See the section “The Graph Template Language” on page 687 and Example 21.5 in Chapter 21, “Statistical Graphics Using ODS,” for more information about dynamic variables and ODS Graphics.

## Examples: LIFETEST Procedure

### Example 49.1: Product-Limit Estimates and Tests of Association

The data presented in Appendix I of Kalbfleisch and Prentice (1980) are coded in the following DATA step. The response variable, SurvTime, is the survival time in days of a lung cancer patient. Negative values of SurvTime are censored values. The covariates are Cell (type of cancer cell), Therapy (type of therapy: standard or test), Prior (prior therapy: 0=no, 1=yes), Age (age in years), DiagTime (time in months from diagnosis to entry into the trial), and Kps (performance status). A censoring indicator variable Censor is created from the data, with the value 1 indicating a censored time and the value 0 indicating an event time. Since there are only two types of therapy, an indicator variable, Treatment, is constructed for therapy type, with value 0 for standard therapy and value 1 for test therapy.

```
data VALung;
  drop check m;
  retain Therapy Cell;
  infile cards column=column;
  length Check $ 1;
  label SurvTime='failure or censoring time'
        Kps='karnofsky index'
        DiagTime='months till randomization'
        Age='age in years'
        Prior='prior treatment?'
        Cell='cell type'
        Therapy='type of treatment'
        Treatment='treatment indicator';
  M=Column;
  input Check $ @@;
  if M>Column then M=1;
  if Check='s'|Check='t' then input @M Therapy $ Cell $ ;
  else input @M SurvTime Kps DiagTime Age Prior @@;
  if SurvTime > .;
  censor=(SurvTime<0);
  SurvTime=abs(SurvTime);
  Treatment=(Therapy='test');
  datalines;
standard squamous
  72 60 7 69 0 411 70 5 64 10 228 60 3 38 0 126 60 9 63 10
118 70 11 65 10 10 20 5 49 0 82 40 10 69 10 110 80 29 68 0
314 50 18 43 0 -100 70 6 70 0 42 60 4 81 0 8 40 58 63 10
144 30 4 63 0 -25 80 9 52 10 11 70 11 48 10
standard small
  30 60 3 61 0 384 60 9 42 0 4 40 2 35 0 54 80 4 63 10
```

```

13 60 4 56 0 -123 40 3 55 0 -97 60 5 67 0 153 60 14 63 10
59 30 2 65 0 117 80 3 46 0 16 30 4 53 10 151 50 12 69 0
22 60 4 68 0 56 80 12 43 10 21 40 2 55 10 18 20 15 42 0
139 80 2 64 0 20 30 5 65 0 31 75 3 65 0 52 70 2 55 0
287 60 25 66 10 18 30 4 60 0 51 60 1 67 0 122 80 28 53 0
27 60 8 62 0 54 70 1 67 0 7 50 7 72 0 63 50 11 48 0
392 40 4 68 0 10 40 23 67 10
standard adeno
8 20 19 61 10 92 70 10 60 0 35 40 6 62 0 117 80 2 38 0
132 80 5 50 0 12 50 4 63 10 162 80 5 64 0 3 30 3 43 0
95 80 4 34 0
standard large
177 50 16 66 10 162 80 5 62 0 216 50 15 52 0 553 70 2 47 0
278 60 12 63 0 12 40 12 68 10 260 80 5 45 0 200 80 12 41 10
156 70 2 66 0 -182 90 2 62 0 143 90 8 60 0 105 80 11 66 0
103 80 5 38 0 250 70 8 53 10 100 60 13 37 10
test squamous
999 90 12 54 10 112 80 6 60 0 -87 80 3 48 0 -231 50 8 52 10
242 50 1 70 0 991 70 7 50 10 111 70 3 62 0 1 20 21 65 10
587 60 3 58 0 389 90 2 62 0 33 30 6 64 0 25 20 36 63 0
357 70 13 58 0 467 90 2 64 0 201 80 28 52 10 1 50 7 35 0
30 70 11 63 0 44 60 13 70 10 283 90 2 51 0 15 50 13 40 10
test small
25 30 2 69 0 -103 70 22 36 10 21 20 4 71 0 13 30 2 62 0
87 60 2 60 0 2 40 36 44 10 20 30 9 54 10 7 20 11 66 0
24 60 8 49 0 99 70 3 72 0 8 80 2 68 0 99 85 4 62 0
61 70 2 71 0 25 70 2 70 0 95 70 1 61 0 80 50 17 71 0
51 30 87 59 10 29 40 8 67 0
test adeno
24 40 2 60 0 18 40 5 69 10 -83 99 3 57 0 31 80 3 39 0
51 60 5 62 0 90 60 22 50 10 52 60 3 43 0 73 60 3 70 0
8 50 5 66 0 36 70 8 61 0 48 10 4 81 0 7 40 4 58 0
140 70 3 63 0 186 90 3 60 0 84 80 4 62 10 19 50 10 42 0
45 40 3 69 0 80 40 4 63 0
test large
52 60 4 45 0 164 70 15 68 10 19 30 4 39 10 53 60 12 66 0
15 30 5 63 0 43 60 11 49 10 340 80 10 64 10 133 75 1 65 0
111 60 5 64 0 231 70 18 67 10 378 80 4 65 0 49 30 3 37 0
;

```

In the following statements, PROC LIFETEST is invoked to compute the product-limit estimate of the survivor function for each type of cancer cell and to analyze the effects of the variables Age, Prior, DiagTime, Kps, and Treatment on the survival of the patients. These prognostic factors are specified in the TEST statement, and the variable Cell is specified in the STRATA statement. Graphics results are enabled through ODS with the specification of the ODS GRAPHICS ON statement. Graphical display of the product-limit estimates (S), the negative log estimates (LS), and the log of negative log estimates (LLS) are requested through the PLOTS= option in the PROC LIFETEST statement. Because of a few large survival times, a MAXTIME of 600 is used to set the scale of the time axis; that is, the time scale extends from 0 to a maximum of 600 days in the plots. The variable Therapy is specified in the ID statement to identify the type of therapy for each observation in the product-limit estimates. The OUTTEST option specifies the creation of an output data set named Test to contain the rank test matrices for the covariates.

```

ods graphics on;
proc lifetest data=VALung plots=(s,ls,lls) outtest=Test maxtime=600;
  time SurvTime*Censor(1);
  id Therapy;
  strata Cell;
  test Age Prior DiagTime Kps Treatment;
run;
ods graphics off;

```

Output 49.1.1 through Output 49.1.4 display the product-limit estimates of the survivor functions for the four cell types. Summary statistics of the survival times are also shown. The median survival times are 51 days, 156 days, 51 days, and 118 days for patients with adeno cells, large cells, small cells, and squamous cells, respectively.

**Output 49.1.1** Estimation Results for Cell=adeno

The LIFETEST Procedure						
Stratum 1: Cell = adeno						
Product-Limit Survival Estimates						
SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	27	
3.000	0.9630	0.0370	0.0363	1	26	standard
7.000	0.9259	0.0741	0.0504	2	25	test
8.000	.	.	.	3	24	standard
8.000	0.8519	0.1481	0.0684	4	23	test
12.000	0.8148	0.1852	0.0748	5	22	standard
18.000	0.7778	0.2222	0.0800	6	21	test
19.000	0.7407	0.2593	0.0843	7	20	test
24.000	0.7037	0.2963	0.0879	8	19	test
31.000	0.6667	0.3333	0.0907	9	18	test
35.000	0.6296	0.3704	0.0929	10	17	standard
36.000	0.5926	0.4074	0.0946	11	16	test
45.000	0.5556	0.4444	0.0956	12	15	test
48.000	0.5185	0.4815	0.0962	13	14	test
51.000	0.4815	0.5185	0.0962	14	13	test
52.000	0.4444	0.5556	0.0956	15	12	test
73.000	0.4074	0.5926	0.0946	16	11	test
80.000	0.3704	0.6296	0.0929	17	10	test
83.000*	.	.	.	17	9	test
84.000	0.3292	0.6708	0.0913	18	8	test
90.000	0.2881	0.7119	0.0887	19	7	test
92.000	0.2469	0.7531	0.0850	20	6	standard
95.000	0.2058	0.7942	0.0802	21	5	standard
117.000	0.1646	0.8354	0.0740	22	4	standard
132.000	0.1235	0.8765	0.0659	23	3	standard
140.000	0.0823	0.9177	0.0553	24	2	test
162.000	0.0412	0.9588	0.0401	25	1	standard
186.000	0	1.0000	.	26	0	test

NOTE: The marked survival times are censored observations.



**Output 49.1.2** Estimation Results for Cell=large

The LIFETEST Procedure						
Stratum 2: Cell = large						
Product-Limit Survival Estimates						
SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	27	
12.000	0.9630	0.0370	0.0363	1	26	standard
15.000	0.9259	0.0741	0.0504	2	25	test
19.000	0.8889	0.1111	0.0605	3	24	test
43.000	0.8519	0.1481	0.0684	4	23	test
49.000	0.8148	0.1852	0.0748	5	22	test
52.000	0.7778	0.2222	0.0800	6	21	test
53.000	0.7407	0.2593	0.0843	7	20	test
100.000	0.7037	0.2963	0.0879	8	19	standard
103.000	0.6667	0.3333	0.0907	9	18	standard
105.000	0.6296	0.3704	0.0929	10	17	standard
111.000	0.5926	0.4074	0.0946	11	16	test
133.000	0.5556	0.4444	0.0956	12	15	test
143.000	0.5185	0.4815	0.0962	13	14	standard
156.000	0.4815	0.5185	0.0962	14	13	standard
162.000	0.4444	0.5556	0.0956	15	12	standard
164.000	0.4074	0.5926	0.0946	16	11	test
177.000	0.3704	0.6296	0.0929	17	10	standard
182.000*	.	.	.	17	9	standard
200.000	0.3292	0.6708	0.0913	18	8	standard
216.000	0.2881	0.7119	0.0887	19	7	standard
231.000	0.2469	0.7531	0.0850	20	6	test
250.000	0.2058	0.7942	0.0802	21	5	standard
260.000	0.1646	0.8354	0.0740	22	4	standard
278.000	0.1235	0.8765	0.0659	23	3	standard
340.000	0.0823	0.9177	0.0553	24	2	test
378.000	0.0412	0.9588	0.0401	25	1	test
553.000	0	1.0000	.	26	0	standard

NOTE: The marked survival times are censored observations.

Output 49.1.3 Estimation Results for Cell=small

The LIFETEST Procedure						
Stratum 3: Cell = small						
Product-Limit Survival Estimates						
SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	48	
2.000	0.9792	0.0208	0.0206	1	47	test
4.000	0.9583	0.0417	0.0288	2	46	standard
7.000	.	.	.	3	45	standard
7.000	0.9167	0.0833	0.0399	4	44	test
8.000	0.8958	0.1042	0.0441	5	43	test
10.000	0.8750	0.1250	0.0477	6	42	standard
13.000	.	.	.	7	41	standard
13.000	0.8333	0.1667	0.0538	8	40	test
16.000	0.8125	0.1875	0.0563	9	39	standard
18.000	.	.	.	10	38	standard
18.000	0.7708	0.2292	0.0607	11	37	standard
20.000	.	.	.	12	36	standard
20.000	0.7292	0.2708	0.0641	13	35	test
21.000	.	.	.	14	34	standard
21.000	0.6875	0.3125	0.0669	15	33	test
22.000	0.6667	0.3333	0.0680	16	32	standard
24.000	0.6458	0.3542	0.0690	17	31	test
25.000	.	.	.	18	30	test
25.000	0.6042	0.3958	0.0706	19	29	test
27.000	0.5833	0.4167	0.0712	20	28	standard
29.000	0.5625	0.4375	0.0716	21	27	test
30.000	0.5417	0.4583	0.0719	22	26	standard
31.000	0.5208	0.4792	0.0721	23	25	standard
51.000	.	.	.	24	24	standard
51.000	0.4792	0.5208	0.0721	25	23	test
52.000	0.4583	0.5417	0.0719	26	22	standard
54.000	.	.	.	27	21	standard
54.000	0.4167	0.5833	0.0712	28	20	standard
56.000	0.3958	0.6042	0.0706	29	19	standard
59.000	0.3750	0.6250	0.0699	30	18	standard
61.000	0.3542	0.6458	0.0690	31	17	test
63.000	0.3333	0.6667	0.0680	32	16	standard
80.000	0.3125	0.6875	0.0669	33	15	test
87.000	0.2917	0.7083	0.0656	34	14	test
95.000	0.2708	0.7292	0.0641	35	13	test
97.000*	.	.	.	35	12	standard
99.000	.	.	.	36	11	test
99.000	0.2257	0.7743	0.0609	37	10	test
103.000*	.	.	.	37	9	test
117.000	0.2006	0.7994	0.0591	38	8	standard
122.000	0.1755	0.8245	0.0567	39	7	standard
123.000*	.	.	.	39	6	standard
139.000	0.1463	0.8537	0.0543	40	5	standard
151.000	0.1170	0.8830	0.0507	41	4	standard
153.000	0.0878	0.9122	0.0457	42	3	standard
287.000	0.0585	0.9415	0.0387	43	2	standard
384.000	0.0293	0.9707	0.0283	44	1	standard
392.000	0	1.0000	.	45	0	standard

NOTE: The marked survival times are censored observations.

**Output 49.1.4** Estimation Results for Cell=squamous

The LIFETEST Procedure						
Stratum 4: Cell = squamous						
Product-Limit Survival Estimates						
SurvTime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	Therapy
0.000	1.0000	0	0	0	35	
1.000	.	.	.	1	34	test
1.000	0.9429	0.0571	0.0392	2	33	test
8.000	0.9143	0.0857	0.0473	3	32	standard
10.000	0.8857	0.1143	0.0538	4	31	standard
11.000	0.8571	0.1429	0.0591	5	30	standard
15.000	0.8286	0.1714	0.0637	6	29	test
25.000	0.8000	0.2000	0.0676	7	28	test
25.000*	.	.	.	7	27	standard
30.000	0.7704	0.2296	0.0713	8	26	test
33.000	0.7407	0.2593	0.0745	9	25	test
42.000	0.7111	0.2889	0.0772	10	24	standard
44.000	0.6815	0.3185	0.0794	11	23	test
72.000	0.6519	0.3481	0.0813	12	22	standard
82.000	0.6222	0.3778	0.0828	13	21	standard
87.000*	.	.	.	13	20	test
100.000*	.	.	.	13	19	standard
110.000	0.5895	0.4105	0.0847	14	18	standard
111.000	0.5567	0.4433	0.0861	15	17	test
112.000	0.5240	0.4760	0.0870	16	16	test
118.000	0.4912	0.5088	0.0875	17	15	standard
126.000	0.4585	0.5415	0.0876	18	14	standard
144.000	0.4257	0.5743	0.0873	19	13	standard
201.000	0.3930	0.6070	0.0865	20	12	test
228.000	0.3602	0.6398	0.0852	21	11	standard
231.000*	.	.	.	21	10	test
242.000	0.3242	0.6758	0.0840	22	9	test
283.000	0.2882	0.7118	0.0820	23	8	test
314.000	0.2522	0.7478	0.0793	24	7	standard
357.000	0.2161	0.7839	0.0757	25	6	test
389.000	0.1801	0.8199	0.0711	26	5	test
411.000	0.1441	0.8559	0.0654	27	4	standard
467.000	0.1081	0.8919	0.0581	28	3	test
587.000	0.0720	0.9280	0.0487	29	2	test
991.000	0.0360	0.9640	0.0352	30	1	test
999.000	0	1.0000	.	31	0	test

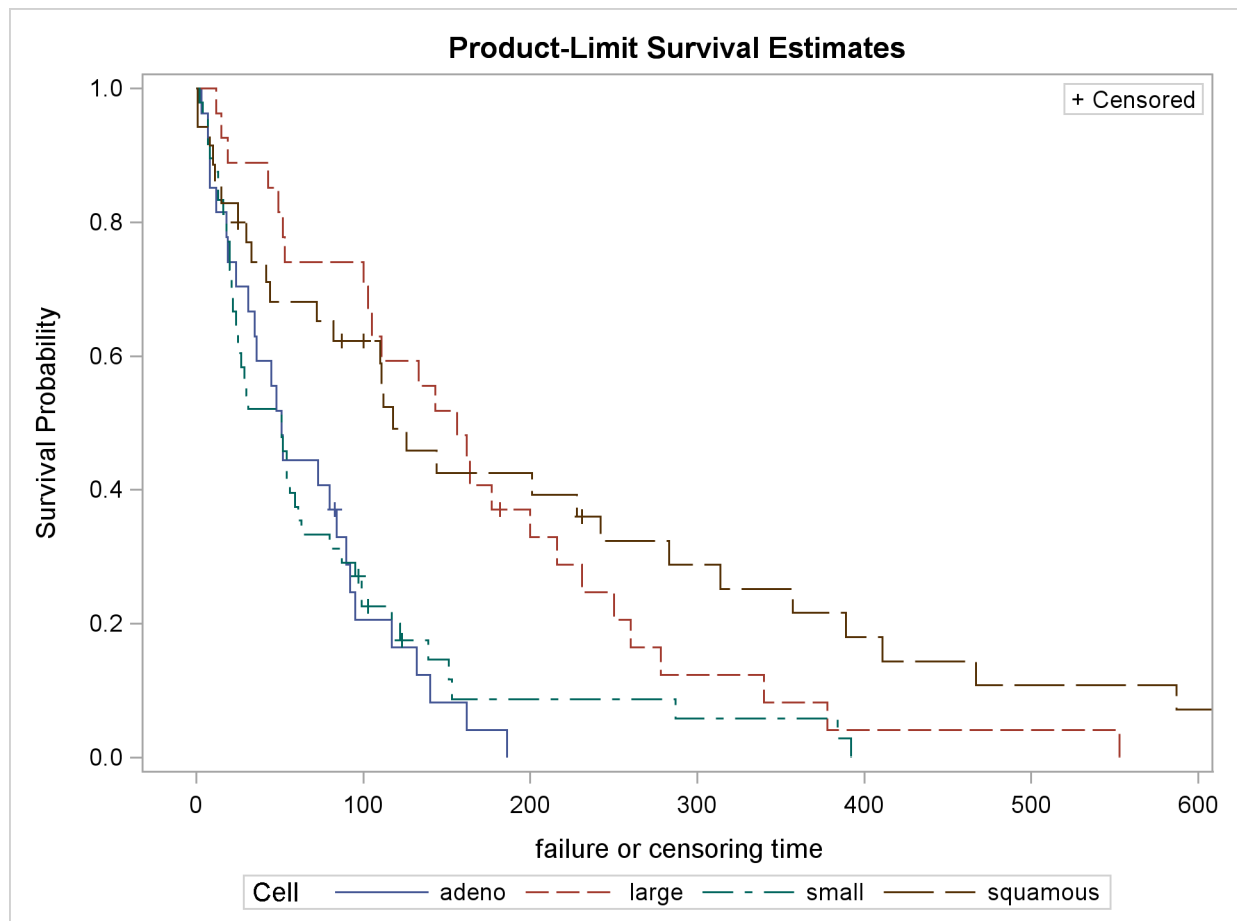
NOTE: The marked survival times are censored observations.

The distribution of event and censored observations among the four cell types is summarized in [Output 49.1.5](#).

**Output 49.1.5** Summary of Censored and Uncensored Values

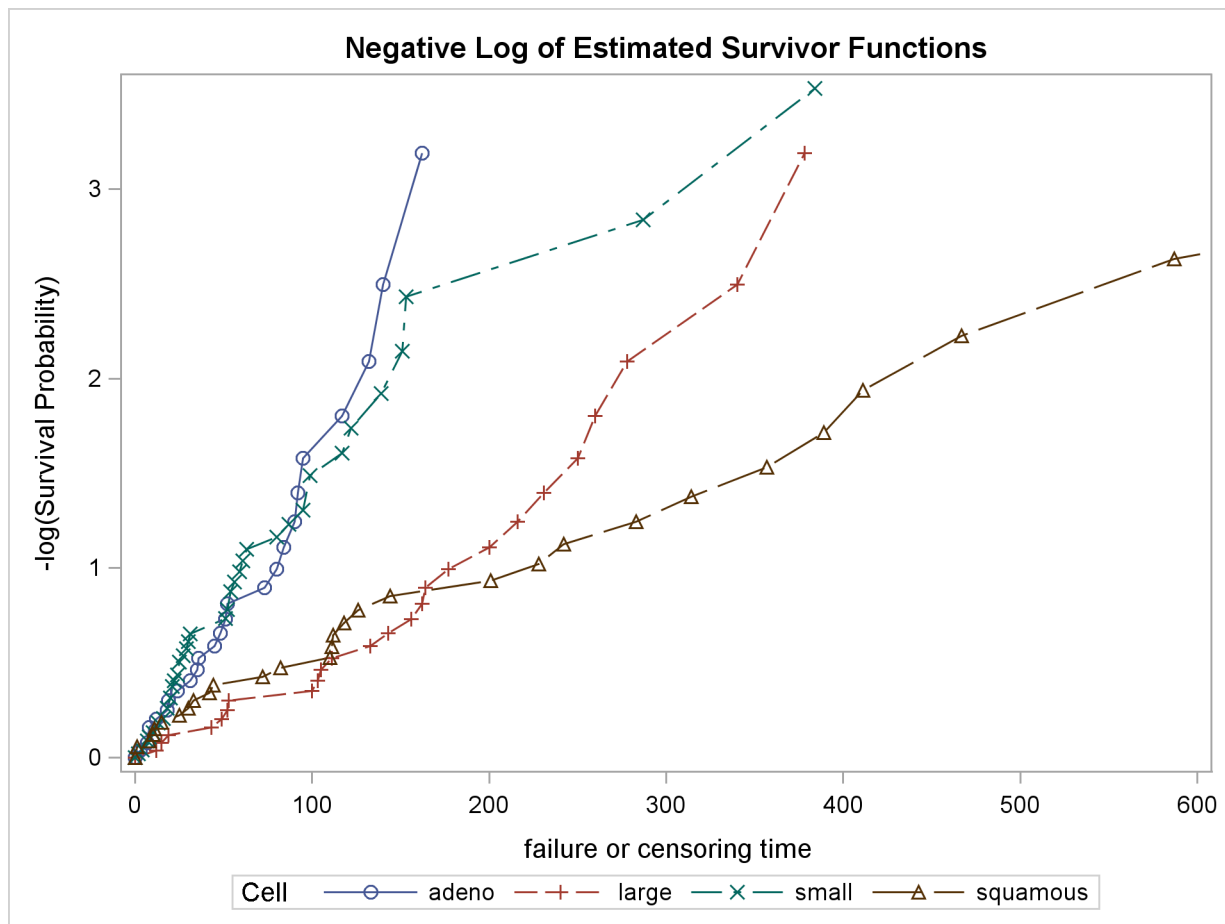
Summary of the Number of Censored and Uncensored Values					
Stratum	Cell	Total	Failed	Censored	Percent Censored
1	adeno	27	26	1	3.70
2	large	27	26	1	3.70
3	small	48	45	3	6.25
4	squamous	35	31	4	11.43
Total		137	128	9	6.57

The graph of the estimated survivor functions is shown in [Output 49.1.6](#). The adeno cell curve and the small cell curve are much closer to each other than they are to the large cell curve or the squamous cell curve. The survival rates of the adeno cell patients and the small cell patients decrease rapidly to approximately 29% in 90 days. Shapes of the large cell curve and the squamous cell curve are quite different, although both decrease less rapidly than those of the adeno and small cells. The squamous cell curve decreases more rapidly initially than the large cell curve, but the role is reversed in the later period.

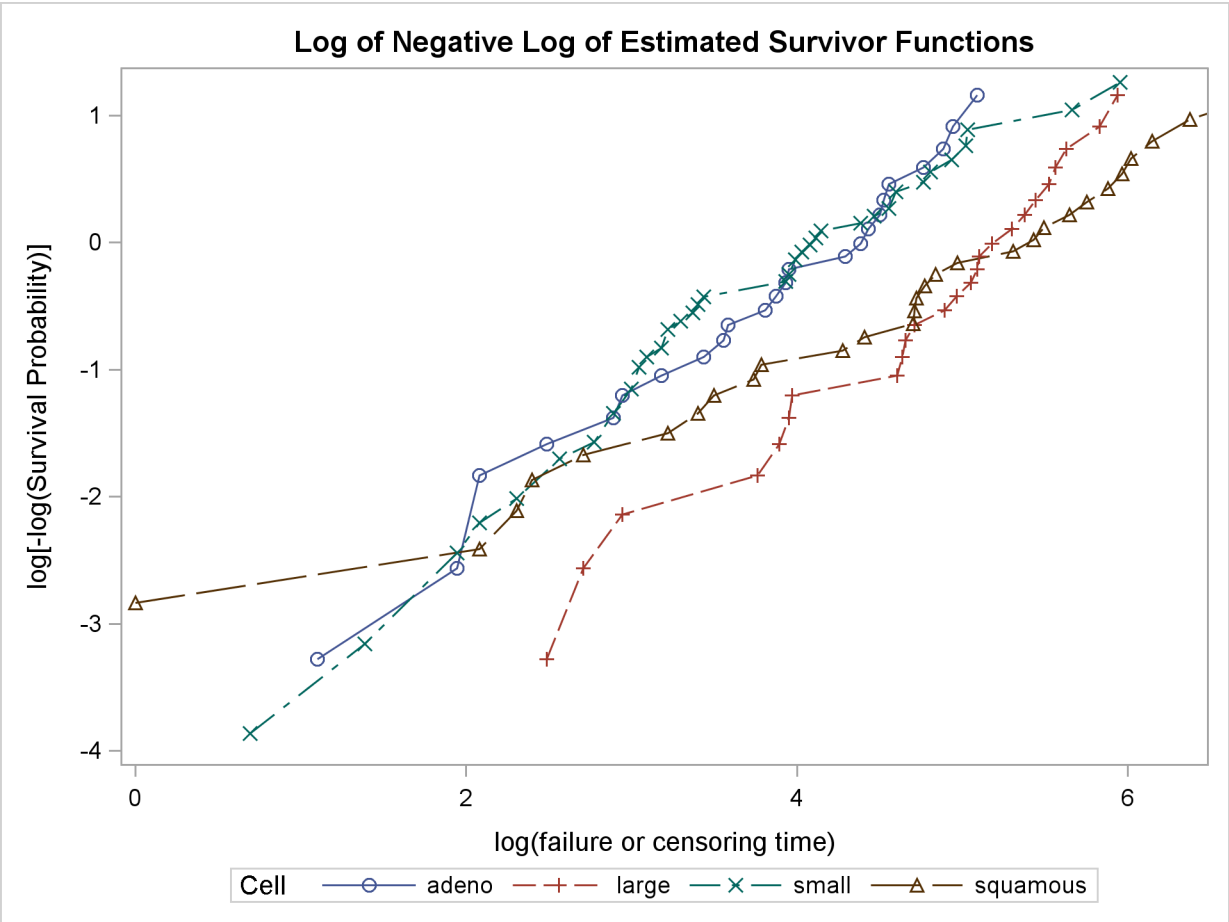
**Output 49.1.6** Graph of the Estimated Survivor Functions

The graph of the negative log of the estimated survivor functions is displayed in [Output 49.1.7](#). [Output 49.1.8](#) displays the log of the negative log of the estimated survivor functions against the log of time.

**Output 49.1.7** Graph of Negative Log of the Estimated Survivor Functions



**Output 49.1.8** Graph of Log of the Negative Log of the Estimated Survivor Functions



Results of the homogeneity tests across cell types are given in [Output 49.1.9](#). The log-rank and Wilcoxon statistics and their corresponding covariance matrices are displayed. Also given is a table that consists of the approximate chi-square statistics, degrees of freedom, and  $p$ -values for the log-rank, Wilcoxon, and likelihood ratio tests. All three tests indicate strong evidence of a significant difference among the survival curves for the four types of cancer cells ( $p < 0.0001$ ).

**Output 49.1.9** Homogeneity Tests across Cell Types

Rank Statistics		
Cell	Log-Rank	Wilcoxon
adeno	10.306	697.0
large	-8.549	-1085.0
small	14.898	1278.0
squamous	-16.655	-890.0

**Output 49.1.9** *continued*

Covariance Matrix for the Log-Rank Statistics				
Cell	adeno	large	small	squamous
adeno	12.9662	-4.0701	-4.4087	-4.4873
large	-4.0701	24.1990	-7.8117	-12.3172
small	-4.4087	-7.8117	21.7543	-9.5339
squamous	-4.4873	-12.3172	-9.5339	26.3384

Covariance Matrix for the Wilcoxon Statistics				
Cell	adeno	large	small	squamous
adeno	121188	-34718	-46639	-39831
large	-34718	151241	-59948	-56576
small	-46639	-59948	175590	-69002
squamous	-39831	-56576	-69002	165410

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	25.4037	3	<.0001
Wilcoxon	19.4331	3	0.0002
-2Log(LR)	33.9343	3	<.0001

Results of the log-rank test of the prognostic variables are shown in [Output 49.1.10](#). The univariate test results correspond to testing each prognostic factor marginally. The joint covariance matrix of these univariate test statistics is also displayed. In computing the overall chi-square statistic, the partial chi-square statistics following a forward stepwise entry approach are tabulated.

Consider the log-rank test in [Output 49.1.10](#). Since the univariate test for Kps has the largest chi-square (43.4747) among all the covariates, Kps is entered first. At this stage, the partial chi-square and the chi-square increment for Kps are the same as the univariate chi-square. Among all the covariates not in the model (Age, Prior, DiagTime, Treatment), Treatment has the largest approximate chi-square increment (1.7261) and is entered next. The approximate chi-square for the model containing Kps and Treatment is  $43.4747 + 1.7261 = 45.2008$  with 2 degrees of freedom. The third covariate entered is Age. The fourth is Prior, and the fifth is DiagTime. The overall chi-square statistic in the last line of the output is the partial chi-square for including all the covariates. It has a value of 46.4200 with 5 degrees of freedom, which is highly significant ( $p < 0.0001$ ).

**Output 49.1.10** Log-Rank Test of the Prognostic Factors

Univariate Chi-Squares for the Log-Rank Test					
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square	Label
Age	-40.7383	105.7	0.1485	0.7000	age in years
Prior	-19.9435	46.9836	0.1802	0.6712	prior treatment?
DiagTime	-115.9	97.8708	1.4013	0.2365	months till randomization
Kps	1123.1	170.3	43.4747	<.0001	karnofsky index
Treatment	-4.2076	5.0407	0.6967	0.4039	treatment indicator

Covariance Matrix for the Log-Rank Statistics					
Variable	Age	Prior	DiagTime	Kps	Treatment
Age	11175.4	-301.2	-892.2	-2948.4	119.3
Prior	-301.2	2207.5	2010.9	78.6	13.9
DiagTime	-892.2	2010.9	9578.7	-2295.3	21.9
Kps	-2948.4	78.6	-2295.3	29015.6	61.9
Treatment	119.3	13.9	21.9	61.9	25.4

Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test					
Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
Kps	1	43.4747	<.0001	43.4747	<.0001
Treatment	2	45.2008	<.0001	1.7261	0.1889
Age	3	46.3012	<.0001	1.1004	0.2942
Prior	4	46.4134	<.0001	0.1122	0.7377
DiagTime	5	46.4200	<.0001	0.00665	0.9350

Variable	Label
Kps	karnofsky index
Treatment	treatment indicator
Age	age in years
Prior	prior treatment?
DiagTime	months till randomization

You can establish this forward stepwise entry of prognostic factors by passing the matrix corresponding to the log-rank test to the RSQUARE method in the REG procedure, as follows. PROC REG finds the sets of variables that yield the largest chi-square statistics.



```

data RSq;
  set Test;
  if _type_='LOG RANK';
  _type_='cov';
proc print data=RSq;
run;
proc reg data=RSq(type=COV);
  model SurvTime=Age Prior DiagTime Kps Treatment
    / selection=rsquare;
  title 'All Possible Subsets of Covariates for the log-rank Test';
run;

```

Output 49.1.11 displays the univariate statistics and their covariance matrix for the log-rank test.

**Output 49.1.11** Log-Rank Statistics and Covariance Matrix

All Possible Subsets of Covariates for the log-rank Test								
Obs	_TYPE_	_NAME_	SurvTime	Age	Prior	DiagTime	Kps	Treatment
1	cov	SurvTime	46.42	-40.74	-19.94	-115.86	1123.14	-4.208
2	cov	Age	-40.74	11175.44	-301.23	-892.24	-2948.45	119.297
3	cov	Prior	-19.94	-301.23	2207.46	2010.85	78.64	13.875
4	cov	DiagTime	-115.86	-892.24	2010.85	9578.69	-2295.32	21.859
5	cov	Kps	1123.14	-2948.45	78.64	-2295.32	29015.62	61.945
6	cov	Treatment	-4.21	119.30	13.87	21.86	61.95	25.409

Results of the best subset regression are shown in [Output 49.1.12](#). The variable Kps generates the largest univariate test statistic among all the covariates, the pair Kps and Age generate the largest test statistic among any other pairs of covariates, and so on. The entry order of covariates is identical to that of PROC LIFETEST.

**Output 49.1.12** Best Subset Regression from the REG Procedure

All Possible Subsets of Covariates for the log-rank Test		
The REG Procedure		
Model: MODEL1		
Dependent Variable: SurvTime		
R-Square Selection Method		
Number in Model	R-Square	Variables in Model
1	0.9366	Kps
1	0.0302	DiagTime
1	0.0150	Treatment
1	0.0039	Prior
1	0.0032	Age
-----		
2	0.9737	Kps Treatment
2	0.9472	Age Kps
2	0.9417	Prior Kps
2	0.9382	DiagTime Kps
2	0.0434	DiagTime Treatment
2	0.0353	Age DiagTime
2	0.0304	Prior DiagTime
2	0.0181	Prior Treatment
2	0.0159	Age Treatment
2	0.0075	Age Prior
-----		
3	0.9974	Age Kps Treatment
3	0.9774	Prior Kps Treatment
3	0.9747	DiagTime Kps Treatment
3	0.9515	Age Prior Kps
3	0.9481	Age DiagTime Kps
3	0.9418	Prior DiagTime Kps
3	0.0456	Age DiagTime Treatment
3	0.0438	Prior DiagTime Treatment
3	0.0355	Age Prior DiagTime
3	0.0192	Age Prior Treatment
-----		
4	0.9999	Age Prior Kps Treatment
4	0.9976	Age DiagTime Kps Treatment
4	0.9774	Prior DiagTime Kps Treatment
4	0.9515	Age Prior DiagTime Kps
4	0.0459	Age Prior DiagTime Treatment
-----		
5	1.0000	Age Prior DiagTime Kps Treatment

## Example 49.2: Enhanced Survival Plot and Multiple-Comparison Adjustments

This example highlights some of the new features of PROC LIFETEST for SAS 9.2, especially the survival plot with number of subjects at risk and multiple comparisons of survival curves. In the following DATA step, data of 137 bone marrow transplant patients extracted from Klein and Moeschberger (1997) are saved in the data set BMT. At the time of transplant, each patient is classified into one of three risk categories: ALL (Acute Lymphoblastic Leukemia), AML (Acute Myelocytic Leukemia)-Low Risk, and AML-High Risk. The endpoint of interest is the disease-free survival time, which is the time to death or relapse or the end of the study in days. In this data set, the variable Group represents the patient's risk category, the variable T represents the disease-free survival time, and the variable Status is the censoring indicator, with the value 1 indicating an event time and the value 0 a censored time.

```
proc format;
  value risk 1='ALL' 2='AML-Low Risk' 3='AML-High Risk';

data BMT;
  input Group T Status @@;
  format Group risk.;
  label T='Disease Free Time';
  datalines;
1 2081 0 1 1602 0 1 1496 0 1 1462 0 1 1433 0
1 1377 0 1 1330 0 1 996 0 1 226 0 1 1199 0
1 1111 0 1 530 0 1 1182 0 1 1167 0 1 418 1
1 383 1 1 276 1 1 104 1 1 609 1 1 172 1
1 487 1 1 662 1 1 194 1 1 230 1 1 526 1
1 122 1 1 129 1 1 74 1 1 122 1 1 86 1
1 466 1 1 192 1 1 109 1 1 55 1 1 1 1
1 107 1 1 110 1 1 332 1 2 2569 0 2 2506 0
2 2409 0 2 2218 0 2 1857 0 2 1829 0 2 1562 0
2 1470 0 2 1363 0 2 1030 0 2 860 0 2 1258 0
2 2246 0 2 1870 0 2 1799 0 2 1709 0 2 1674 0
2 1568 0 2 1527 0 2 1324 0 2 957 0 2 932 0
2 847 0 2 848 0 2 1850 0 2 1843 0 2 1535 0
2 1447 0 2 1384 0 2 414 1 2 2204 1 2 1063 1
2 481 1 2 105 1 2 641 1 2 390 1 2 288 1
2 421 1 2 79 1 2 748 1 2 486 1 2 48 1
2 272 1 2 1074 1 2 381 1 2 10 1 2 53 1
2 80 1 2 35 1 2 248 1 2 704 1 2 211 1
2 219 1 2 606 1 3 2640 0 3 2430 0 3 2252 0
3 2140 0 3 2133 0 3 1238 0 3 1631 0 3 2024 0
3 1345 0 3 1136 0 3 845 0 3 422 1 3 162 1
3 84 1 3 100 1 3 2 1 3 47 1 3 242 1
3 456 1 3 268 1 3 318 1 3 32 1 3 467 1
3 47 1 3 390 1 3 183 1 3 105 1 3 115 1
3 164 1 3 93 1 3 120 1 3 80 1 3 677 1
3 64 1 3 168 1 3 74 1 3 16 1 3 157 1
3 625 1 3 48 1 3 273 1 3 63 1 3 76 1
3 113 1 3 363 1
;
```

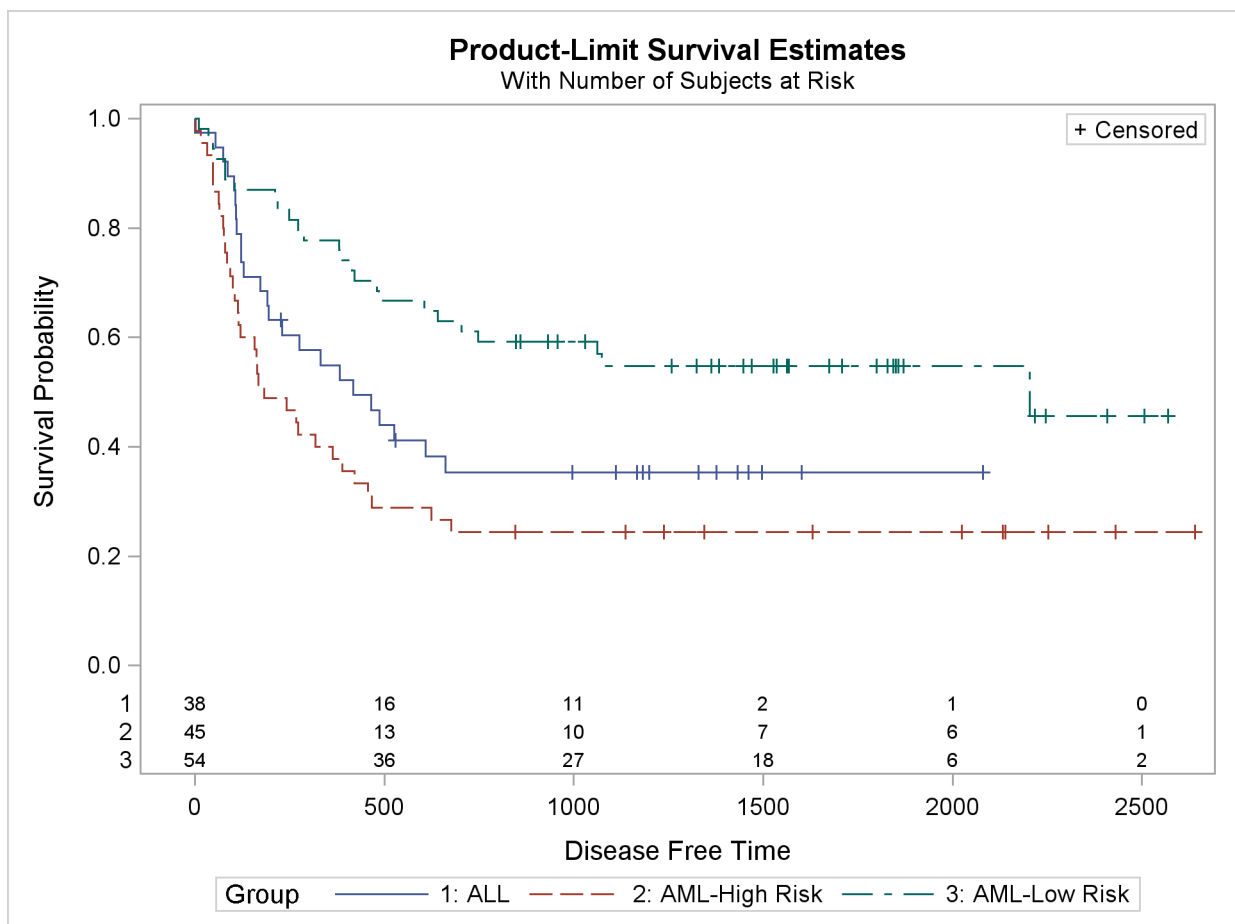
In the following statements, PROC LIFETEST is invoked to compute the product-limit estimate of the survivor function for each risk category. Using ODS Graphics, you can display the number of subjects at risk in the survival plot. The **PLOTS=** option requests that the survival curves be plotted, and the **ATRISK=** suboption specifies the time points at which the at-risk numbers are displayed. In the STRATA statement, the **ADJUST=SIDAK** option requests the Šidák multiple-comparison adjustment, and by default, all paired comparisons are carried out.

```
ods graphics on;
```

```
proc lifetest data=BMT plots=survival(atrisk=0 to 2500 by 500);
  time T * Status(0);
  strata Group / test=logrank adjust=sidak;
run;
```

Output 49.2.1 displays the estimated disease free survival for the three leukemia groups with the number of subjects at risk at 0, 500, 1000, 1500, 2000, and 2500 days. Patients in the AML-Low Risk group experience disease free longer than those in the ALL group, who in turn fare better than those in the AML-High Risk group.

**Output 49.2.1** Estimated Disease-Free Survival for 137 Bone Marrow Transplant Patients



The log-rank test (Output 49.2.2) shows that the disease-free survival times for these three risk groups are significantly different ( $p=0.001$ ).

**Output 49.2.2** Log-Rank Test of Disease Group Homogeneity

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	13.8037	2	0.0010

The Šidák multiple-comparison results are shown in [Output 49.2.3](#). There is no significant difference in disease-free survivor functions between the ALL and AML-High Risk groups ( $p=0.2779$ ). The difference between the ALL and AML-Low Risk groups is marginal ( $p=0.0685$ ), but the AML-Low Risk and AML-High Risk groups have significantly different disease-free survivor functions ( $p=0.0006$ ).

**Output 49.2.3** All Paired Comparisons

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison		Chi-Square	p-Values	
Group	Group		Raw	Sidak
ALL	AML-High Risk	2.6610	0.1028	0.2779
ALL	AML-Low Risk	5.1400	0.0234	0.0685
AML-High Risk	AML-Low Risk	13.8011	0.0002	0.0006

Suppose you consider the AML-Low Risk group as the reference group. You can use the DIFF= option in the STRATA statement to designate this risk group as the control and apply a multiple-comparison adjustment to the  $p$ -values for the paired comparison between the AML-Low Risk group with each of the other groups. Consider the Šidák correction again. You specify the ADJUST= and DIFF= options as in the following statements.

```
proc lifetest data=BMT notable plots=none;
  time T * Status(0);
  strata Group / test=logrank adjust=sidak diff=control('AML-Low Risk');
run;
```

[Output 49.2.4](#) shows that although both the ALL and AML-High Risk groups differ from the AML-Low Risk group at the 0.05 level, the difference between the AML-High Risk and the AML-Low Risk group is highly significant ( $p=0.0004$ ).

**Output 49.2.4** Comparisons with the Reference Group

The LIFETEST Procedure				
Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison		Chi-Square	p-Values	
Group	Group		Raw	Sidak
ALL	AML-Low Risk	5.1400	0.0234	0.0462
AML-High Risk	AML-Low Risk	13.8011	0.0002	0.0004

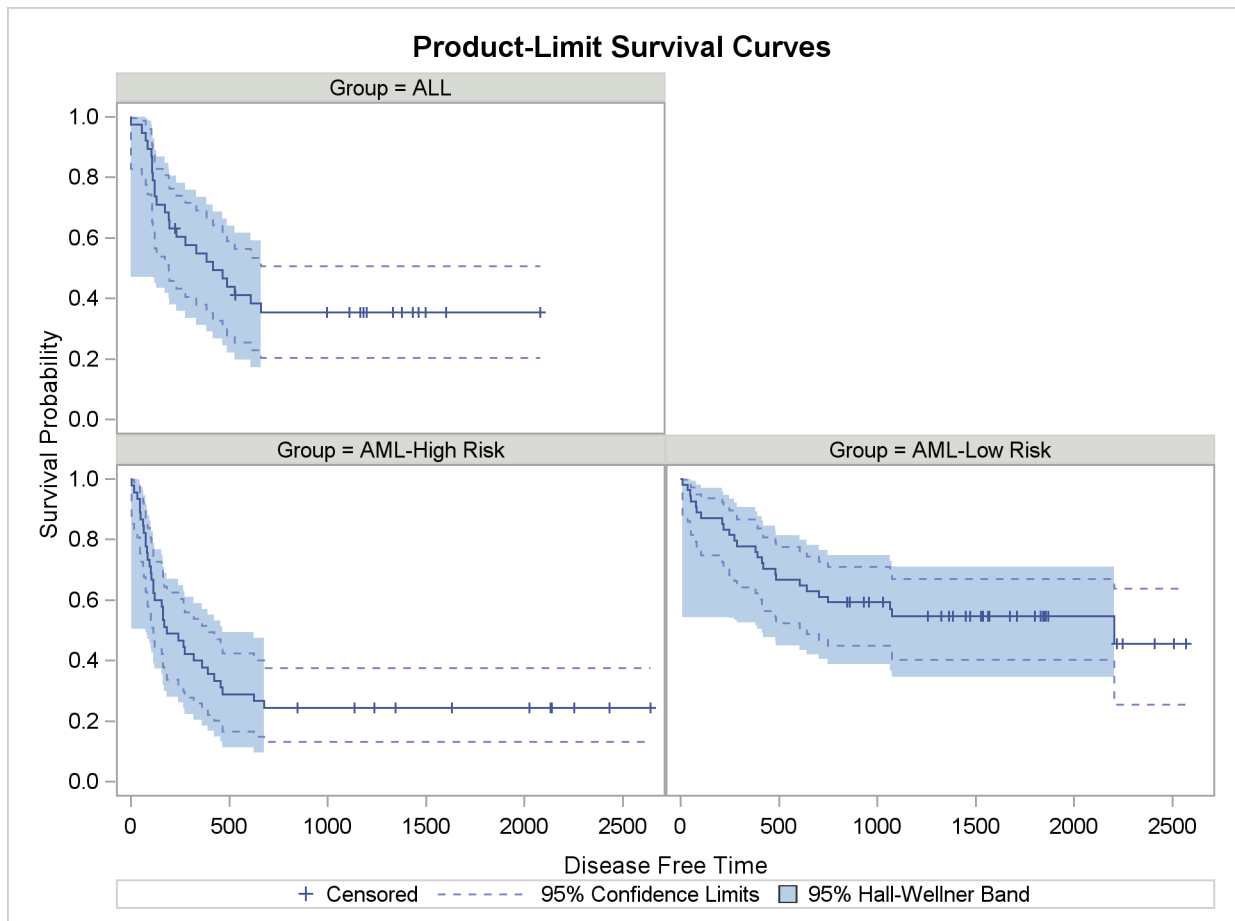
Klein and Moeschberger (1997, Section 4.4) describe in detail how to compute the Hall-Wellner (HW) and equal-precision (EP) confidence bands for the survivor function. You can output these simultaneous confidence intervals to a SAS data set by using the CONFBAND= and OUTSURV= options in the PROC LIFETEST statement. You can display survival curves with pointwise and simultaneous confidence limits through ODS Graphics. When the survival data are stratified, displaying all the survival curves and their confidence limits in the same plot can make the plot appear cluttered. In the following statements, the PLOTS= specification requests that the survivor functions be displayed along with their pointwise confidence limits (CL) and Hall-Wellner confidence bands (CB=HW). The STRATA= PANEL specification requests that the survival curves be displayed in a panel of three plots, one for each risk group.

```
proc lifetest data=BMT plots=survival(cl cb=hw strata=panel);
  time T * Status(0);
  strata Group;
run;

ods graphics off;
```

The panel plot is shown in [Output 49.2.5](#).

**Output 49.2.5** Estimated Disease-Free Survivor Functions with Confidence Limits



### Example 49.3: Life-Table Estimates for Males with Angina Pectoris

The data in this example come from Lee (1992, p. 91) and represent the survival rates of males with angina pectoris. Survival time is measured as years from the time of diagnosis. In the following DATA step, the data are read as number of events and number of withdrawals in each one-year time interval for 16 intervals. Three variables are constructed from the data: Years (an artificial time variable with values that are the midpoints of the time intervals), Censored (a censoring indicator variable with the value 1 indicating censored observations and the value 0 indicating event observations), and Freq (the frequency variable). Two observations are created for each interval, one representing the event observations and the other representing the censored observations.

```

title 'Survival of Males with Angina Pectoris';
data Males;
  keep Freq Years Censored;
  retain Years -.5;
  input fail withdraw @@;
  Years + 1;
  Censored=0;
  Freq=fail;
  output;
  Censored=1;
  Freq=withdraw;
  output;
  datalines;
456  0 226  39 152  22 171  23 135  24 125 107
 83 133  74 102  51  68  42  64  43 45  34  53
 18  33   9  27   6  23   0  30
;

```

In the following statements, the ODS GRAPHICS ON specification enables ODS Graphics. PROC LIFETEST is invoked to compute the various life-table survival estimates, the median residual time, and their standard errors. The life-table method of computing estimates is requested by specifying METHOD=LT. The intervals are specified by the INTERVAL= option. Graphical display of the life-table survivor function estimate, negative log of the estimate, log of negative log of the estimate, estimated density function, and estimated hazard function are requested by the PLOTS= option. No tests for homogeneity are carried out because the data are not stratified.

```

ods graphics on;
proc lifetest data=Males method=lt intervals=(0 to 15 by 1)
  plots=(s,ls,lls,h,p);
  time Years*Censored(1);
  freq Freq;
run;
ods graphics off;

```

Results of the life-table estimation are shown in [Output 49.3.1](#). The five-year survival rate is 0.5193 with a standard error of 0.0103. The estimated median residual lifetime, which is 5.33 years initially, reaches a maximum of 6.34 years at the beginning of the second year and decreases gradually to a value lower than the initial 5.33 years at the beginning of the seventh year.

**Output 49.3.1** Life-Table Survivor Function Estimate

Survival of Males with Angina Pectoris						
The LIFETEST Procedure						
Life Table Survival Estimates						
Interval [Lower, Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	
0	1	456	0	2418.0	0.1886	0.00796
1	2	226	39	1942.5	0.1163	0.00728
2	3	152	22	1686.0	0.0902	0.00698
3	4	171	23	1511.5	0.1131	0.00815
4	5	135	24	1317.0	0.1025	0.00836
5	6	125	107	1116.5	0.1120	0.00944
6	7	83	133	871.5	0.0952	0.00994
7	8	74	102	671.0	0.1103	0.0121
8	9	51	68	512.0	0.0996	0.0132
9	10	42	64	395.0	0.1063	0.0155
10	11	43	45	298.5	0.1441	0.0203
11	12	34	53	206.5	0.1646	0.0258
12	13	18	33	129.5	0.1390	0.0304
13	14	9	27	81.5	0.1104	0.0347
14	15	6	23	47.5	0.1263	0.0482
15	.	0	30	15.0	0	0

Interval [Lower, Upper)	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error
0	1	1.0000	0	5.3313	0.1749
1	2	0.8114	0.1886	6.2499	0.2001
2	3	0.7170	0.2830	6.3432	0.2361
3	4	0.6524	0.3476	6.2262	0.2361
4	5	0.5786	0.4214	6.2185	0.1853
5	6	0.5193	0.4807	5.9077	0.1806
6	7	0.4611	0.5389	5.5962	0.1855
7	8	0.4172	0.5828	5.1671	0.2713
8	9	0.3712	0.6288	4.9421	0.2763
9	10	0.3342	0.6658	4.8258	0.4141
10	11	0.2987	0.7013	4.6888	0.4183
11	12	0.2557	0.7443	.	.
12	13	0.2136	0.7864	.	.
13	14	0.1839	0.8161	.	.
14	15	0.1636	0.8364	.	.



**Output 49.3.1** *continued*

Survival of Males with Angina Pectoris						
The LIFETEST Procedure						
Interval [Lower, Upper)		Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error
15	.	0.1429	0.8571	0.0133	.	.
Evaluated at the Midpoint of the Interval						
Interval [Lower, Upper)		PDF	PDF Standard Error	Hazard	Hazard Standard Error	
0	1	0.1886	0.00796	0.208219	0.009698	
1	2	0.0944	0.00598	0.123531	0.008201	
2	3	0.0646	0.00507	0.09441	0.007649	
3	4	0.0738	0.00543	0.119916	0.009154	
4	5	0.0593	0.00495	0.108043	0.009285	
5	6	0.0581	0.00503	0.118596	0.010589	
6	7	0.0439	0.00469	0.1	0.010963	
7	8	0.0460	0.00518	0.116719	0.013545	
8	9	0.0370	0.00502	0.10483	0.014659	
9	10	0.0355	0.00531	0.112299	0.017301	
10	11	0.0430	0.00627	0.155235	0.023602	
11	12	0.0421	0.00685	0.17942	0.030646	
12	13	0.0297	0.00668	0.149378	0.03511	
13	14	0.0203	0.00651	0.116883	0.038894	
14	15	0.0207	0.00804	0.134831	0.054919	
15	.	.	.	.	.	.

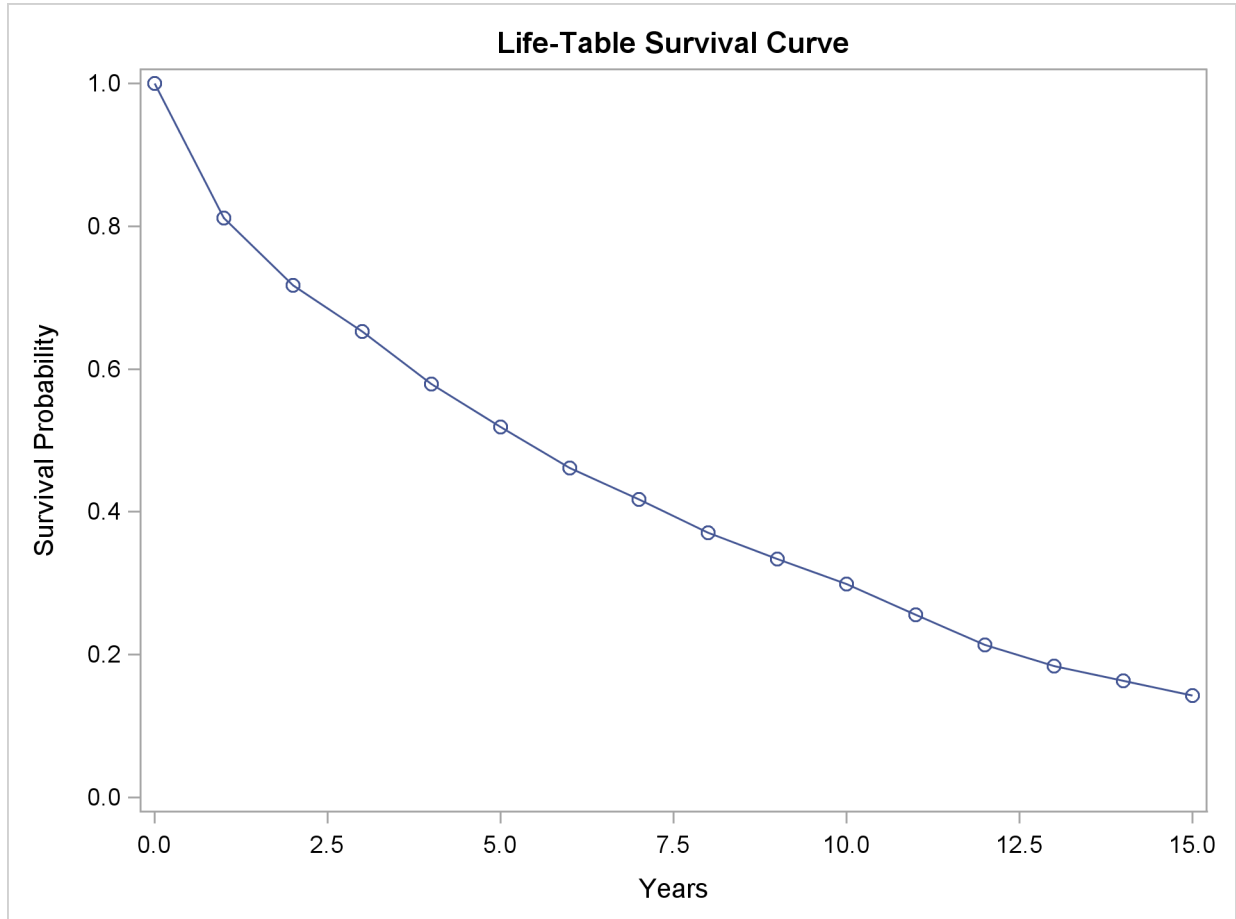
The breakdown of event and censored observation in the data is shown in [Output 49.3.2](#). Note that 32.8% of the patients have withdrawn from the study.

**Output 49.3.2** Summary of Censored and Event Observations

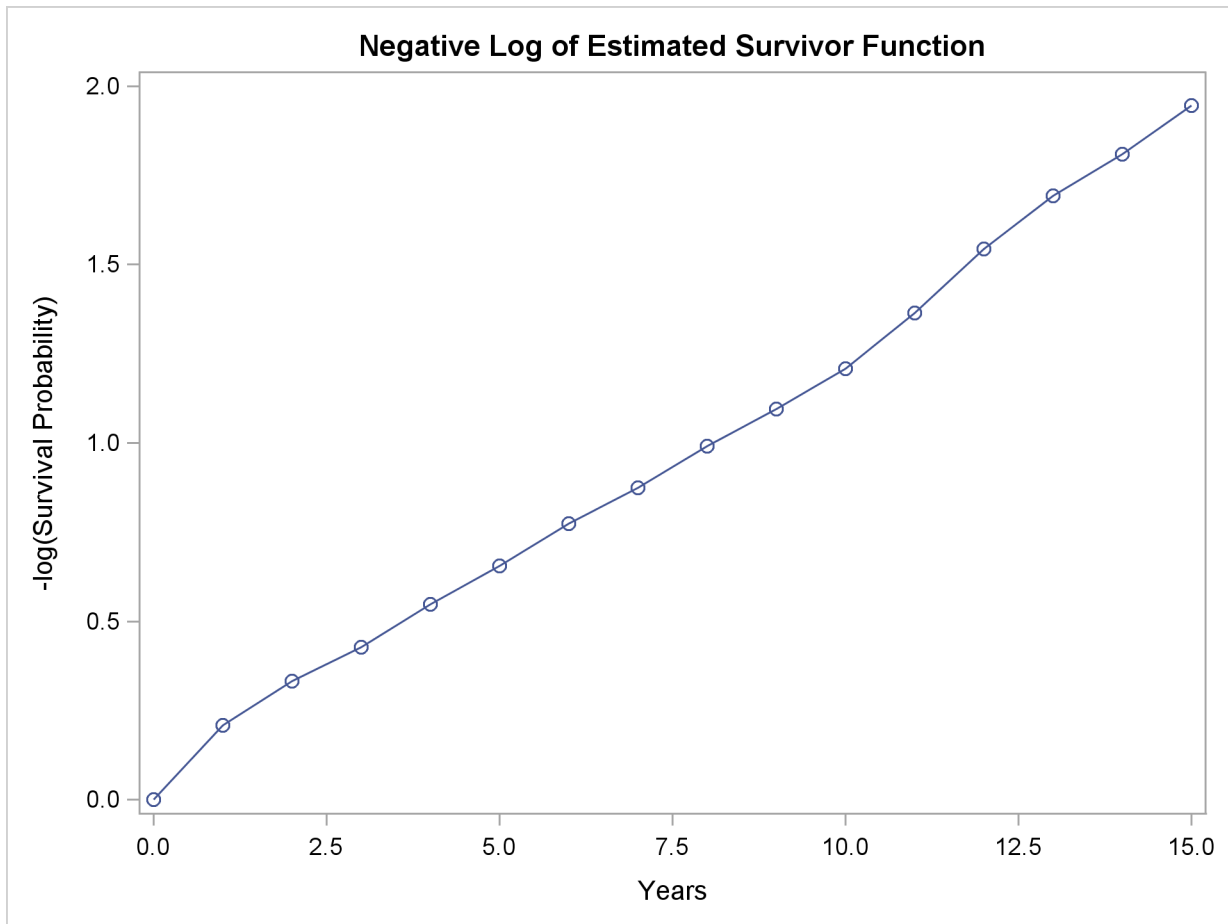
Summary of the Number of Censored and Uncensored Values				
	Total	Failed	Censored	Percent Censored
	2418	1625	793	32.80
NOTE: 2 observations with invalid time, censoring, or frequency values were deleted.				

**Output 49.3.3** displays the graph of the life-table survivor function estimate. The median survival time, read from the survivor function curve, is 5.33 years, and the 25th and 75th percentiles are 1.04 and 11.13 years, respectively.

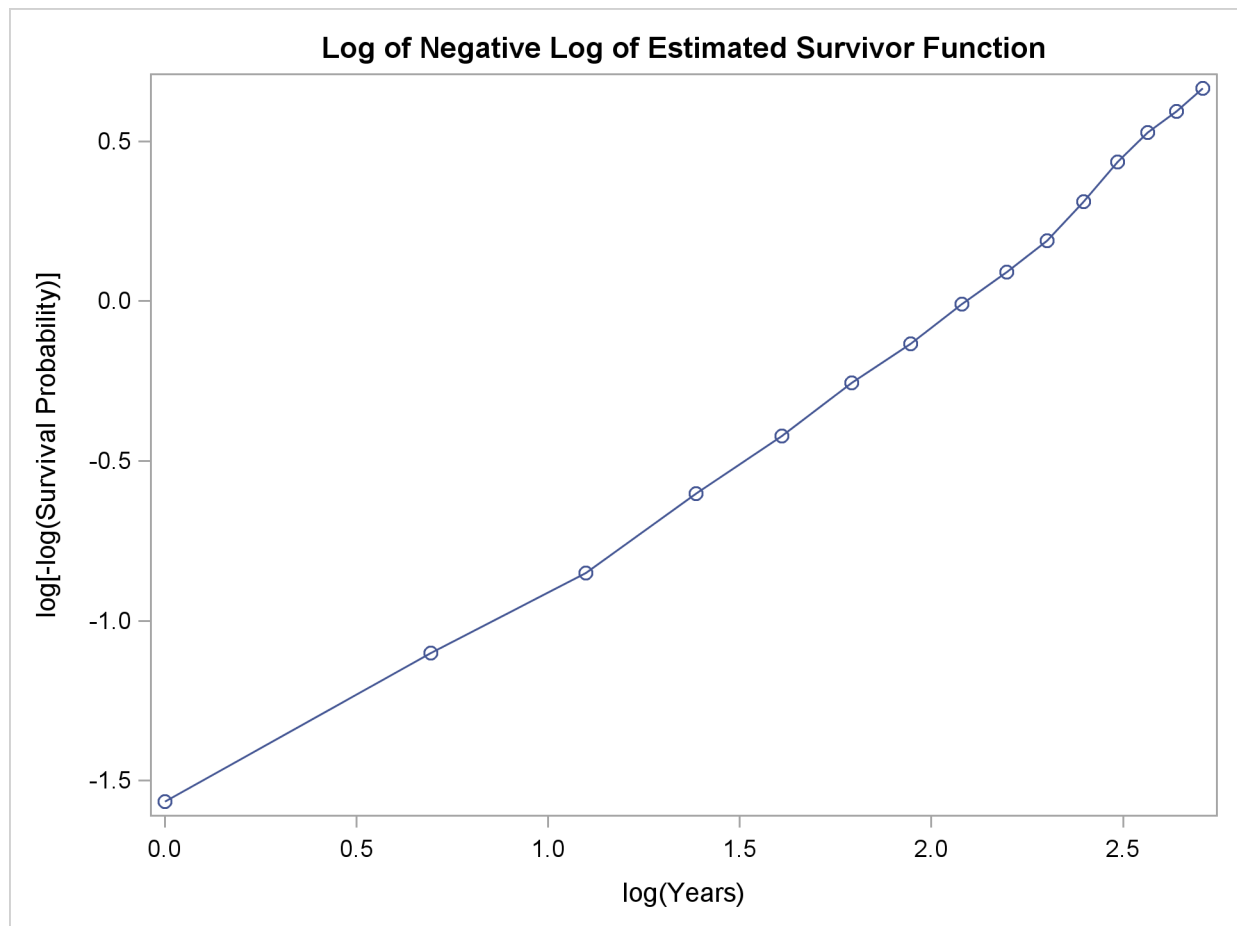
**Output 49.3.3** Life-Table Survivor Function Estimate



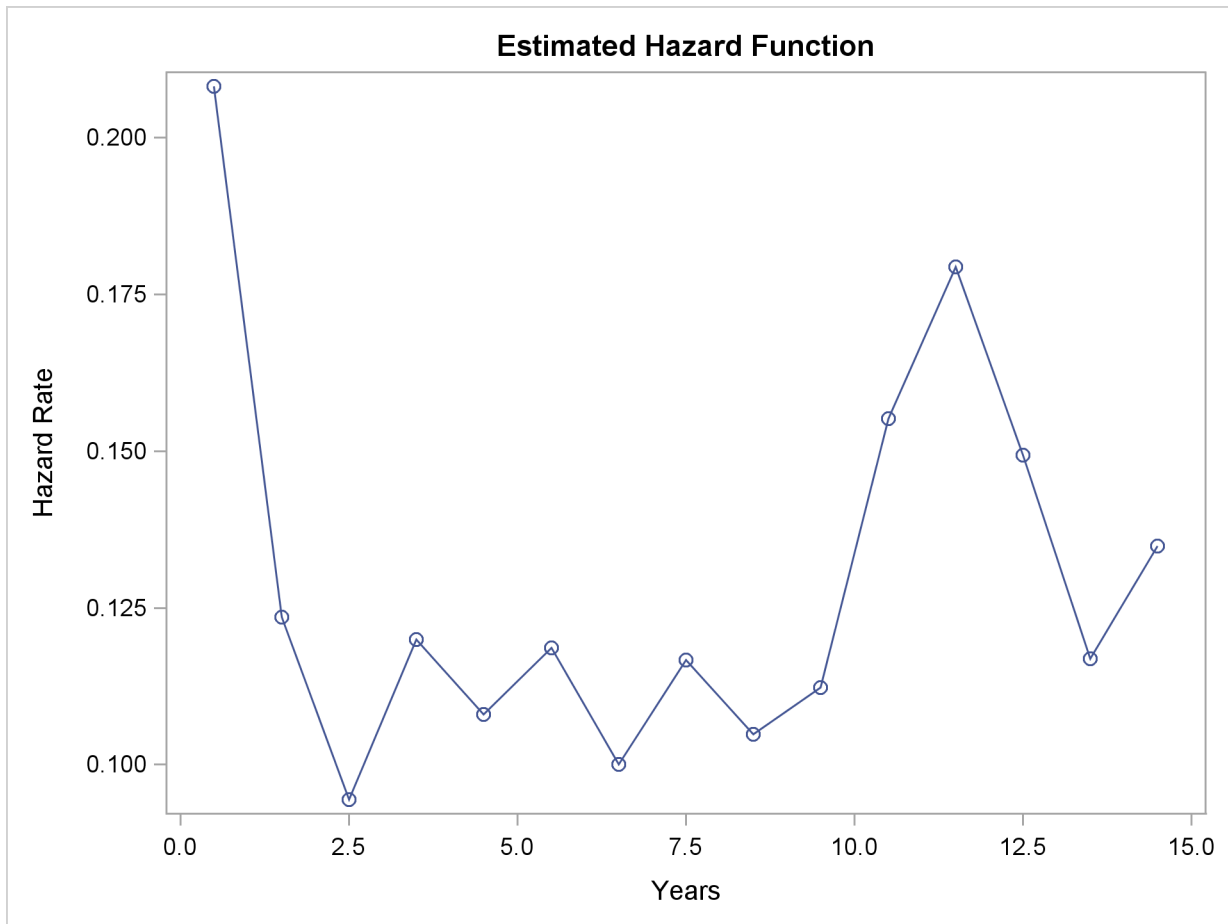
An exponential model might be appropriate for the survival of these male patients with angina pectoris since the curve of the negative log of the survivor function estimate versus the survival time (**Output 49.3.4**) approximates a straight line through the origin. Note that the graph of the log of the negative log of the survivor function estimate versus the log of time (**Output 49.3.5**) is practically a straight line.

**Output 49.3.4** Negative Log of Survivor Function Estimate

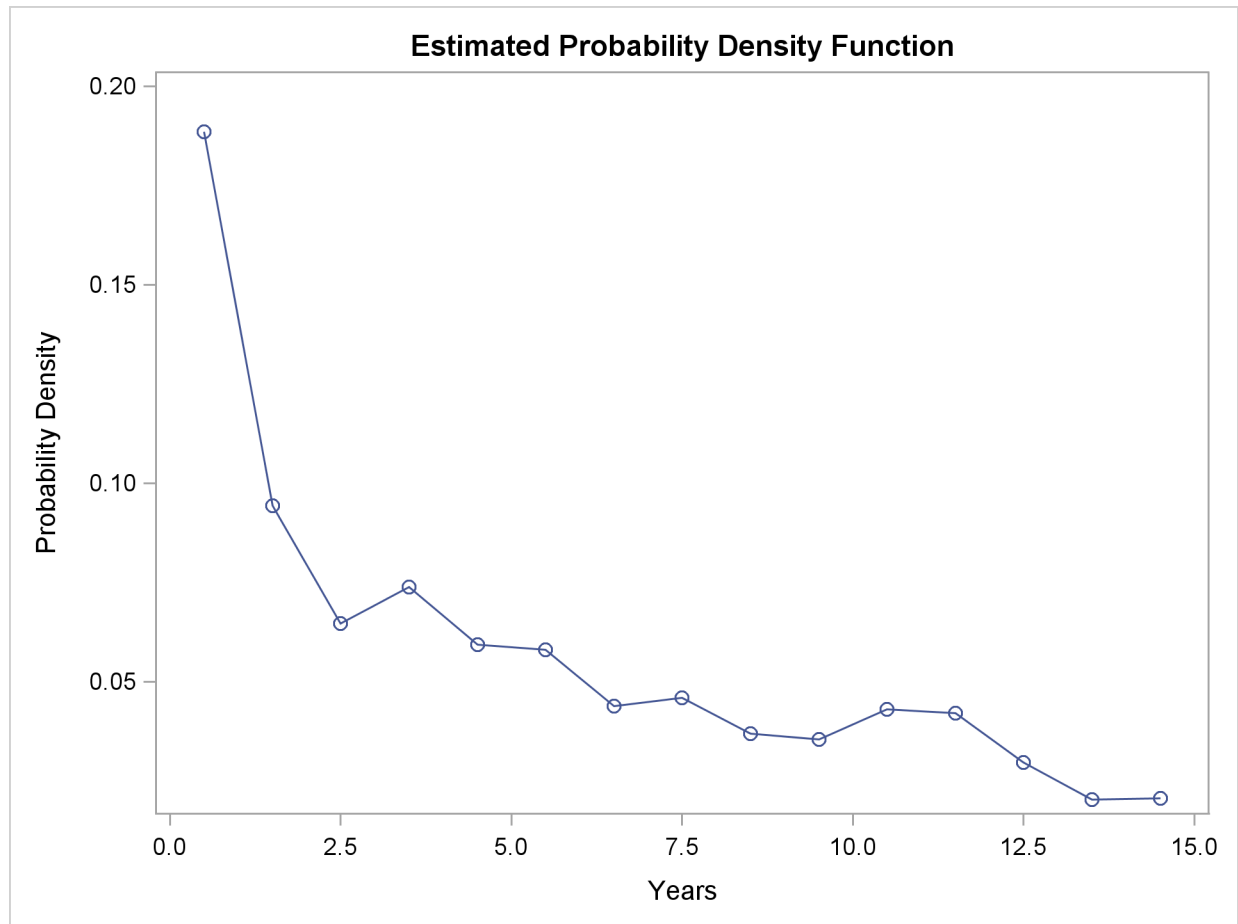
As discussed in Lee (1992), the graph of the estimated hazard function ([Output 49.3.6](#)) shows that the death rate is highest in the first year of diagnosis. From the end of the first year to the end of the tenth year, the death rate remains relatively constant, fluctuating between 0.09 and 0.12. The death rate is generally higher after the tenth year. This could indicate that a patient who has survived the first year has a better chance than a patient who has just been diagnosed. The profile of the median residual lifetimes also supports this interpretation.

**Output 49.3.5** Log of Negative Log of Survivor Function Estimate

**Output 49.3.6** Hazard Function Estimate



The density estimate is shown in (Output 49.3.7). Visually, it resembles the density function of an exponential distribution.

**Output 49.3.7** Density Function Estimate

---

## References

- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1992), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.
- Borgan, Ø. and Liestøl, K. (1990), "A Note on Confidence Interval and Bands for the Survival Curves Based on Transformations," *Scandinavian Journal of Statistics*, 18, 35–41.
- Brookmeyer, R. and Crowley, J. (1982), "A Confidence Interval for the Median Survival Time," *Biometrics*, 38, 29–41.
- Chung, C. F. (1986), *Formulae for Probabilities Associated with Wiener and Brownian Bridge Processes*, Technical Report 79, Laboratory for Research in Statistics and Probability, Ottawa, Canada: Carleton University.
- Collett, D. (1994), *Modeling Survival Data in Medical Research*, London: Chapman & Hall.
- Cox, D. R. and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman & Hall.

- Edwards, D. and Berry, J. J. (1987), "The Efficiency of Simulation-Based Multiple Comparisons," *Biometrics*, 43, 913–928.
- Elandt-Johnson, R. C. and Johnson, N. L. (1980), *Survival Models and Data Analysis*, New York: John Wiley & Sons.
- Fleming, T. R. and Harrington, D. (1991), *Counting Processes and Survival Analysis*, New York: John Wiley & Sons.
- Fleming, T. R. and Harrington, D. P. (1981), "A Class of Hypothesis Tests for One and Two Samples of Censored Survival Data," *Communications in Statistics*, 10, 763–794.
- Fleming, T. R. and Harrington, D. P. (1984), "Nonparametric Estimation of the Survival Distribution in Censored Data," *Communications in Statistics—Theory and Methods*, 13, 2469–2486.
- Gasser, T. and Müller, H. G. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics 757*, 23–68, Berlin: Springer-Verlag.
- Hall, W. J. and Wellner, J. A. (1980), "Confidence Bands for a Survival Curve for Censored Data," *Biometrika* 69.
- Harrington, D. P. and Fleming, T. R. (1982), "A Class of Rank Test Procedures for Censored Survival Data," *Biometrika*, 69, 133–143.
- Hsu, J. C. (1992), "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model," *Journal of Computational and Graphical Statistics*, 1, 151–168.
- Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.
- Klein, J. P. and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer-Verlag.
- Kramer, C. Y. (1956), "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications," *Biometrics*, 12, 307–310.
- Lachin, J. M. (2000), *Biostatistical Methods: The Assessment of Relative Risks*, New York: John Wiley & Sons.
- Lawless, J. F. (1982), *Statistical Methods and Methods for Lifetime Data*, New York: John Wiley & Sons.
- Lee, E. T. (1992), *Statistical Methods for Survival Data Analysis*, Second Edition, New York: John Wiley & Sons.
- Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley & Sons.
- Miller, R. G. and Siegmund, D. (1982), "Maximally Selected Chi-Square Statistics," *Biometrics*, 1011–1016.
- Nair, V. N. (1984), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," *Technometrics*, 26, 265–275.

Ramlau-Hansen, H. (1983a), “The Choice of a Kernel Function in the Graduation of Counting Process Intensities,” *Scandinavian Actuarial Journal*, 165–182.

Ramlau-Hansen, H. (1983b), “Smoothing Counting Process Intensities by Means of Kernel Functions,” *Annual of Statistics*, 11, 453–466.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., and Hochberg, Y. (1999), *Multiple Comparisons and Multiple Tests Using the SAS System*, Cary, NC: SAS Institute Inc.



# Subject Index

- actuarial estimates, *see* life-table estimates
- alpha level
  - LIFETEST procedure, 3716, 3717
- annotate
  - traditional graphics (LIFETEST), 3717
- arcsine-square root transformation
  - confidence intervals (LIFETEST), 3718, 3741, 3743
- association tests
  - LIFETEST procedure, 3704, 3705, 3712, 3775
- at-risk
  - product-limit estimates (LIFETEST), 3717
- Bonferroni adjustment
  - LIFETEST procedure, 3731
- Breslow estimates
  - LIFETEST procedure, 3704, 3736
- Breslow test, *see* Wilcoxon test for homogeneity
- catalog
  - traditional graphics (LIFETEST), 3719
- CDF, *see* cumulative distribution function
- censored
  - LIFETEST procedure, 3703, 3735
- censored symbol
  - traditional graphics (LIFETEST), 3717
- character set
  - line printer plots (LIFETEST), 3719
- confidence bands
  - LIFETEST procedure, 3717, 3742
- confidence limits
  - LIFETEST procedure, 3740, 3753, 3754
- cumulative distribution function
  - LIFETEST procedure, 3704
- density function, *see* probability density function
- description
  - traditional graphics (LIFETEST), 3718
- Dunnett's adjustment
  - LIFETEST procedure, 3731
- effective sample size
  - LIFETEST procedure, 3739
- equal-precision bands
  - LIFETEST procedure, 3718, 3744, 3782
- event symbol
  - traditional graphics (LIFETEST), 3719
- Fleming-Harrington  $G_\rho$  test for homogeneity
  - LIFETEST procedure, 3704, 3734
- Fleming-Harrington estimates
  - LIFETEST procedure, 3704, 3736
- Gehan test, *see* Wilcoxon test for homogeneity
- Hall-Wellner bands
  - LIFETEST procedure, 3718, 3743, 3782
- hazard function
  - LIFETEST procedure, 3704, 3789
- homogeneity tests
  - LIFETEST procedure, 3704, 3711, 3746, 3774
- interval determination
  - LIFETEST procedure, 3740
- interval width
  - life-table method (LIFETEST), 3728
- intervals
  - life-table estimates (LIFETEST), 3719
- k-sample tests, *see* homogeneity tests
- Kaplan-Meier estimates, *see* product-limit estimates
- kernel-smoothed hazard
  - LIFETEST procedure, 3723, 3744
- life-table estimates
  - LIFETEST procedure, 3704, 3758, 3786
- LIFETEST procedure
  - alpha level, 3716, 3717
  - association tests, 3704, 3705, 3712, 3750, 3766, 3775
  - Bonferroni adjustment, 3731
  - Breslow estimates, 3704, 3720, 3736
  - censored, 3703, 3735
  - computational formulas, 3736
  - confidence bands, 3717, 3742
  - confidence limits, 3740, 3753, 3754
  - cumulative distribution function, 3704
  - Dunnett's adjustment, 3731
  - effective sample size, 3739
  - equal-precision bands, 3744, 3782
  - estimation method, 3720
  - Fleming-Harrington  $G_\rho$  test for homogeneity, 3704, 3734
  - Fleming-Harrington estimates, 3704, 3720, 3736

- Hall-Wellner bands, 3743, 3782
- hazard function, 3704, 3789
- homogeneity tests, 3704, 3711, 3746, 3774
- input data set, 3718
- interval determination, 3740
- kernel-smoothed hazard, 3723, 3744
- life-table estimates, 3704, 3720, 3739, 3758, 3783, 3786
- likelihood ratio test for homogeneity, 3704, 3746
- line printer plots, 3715
- log-rank test for association, 3705, 3750
- log-rank test for homogeneity, 3704, 3734, 3746
- maximum time, 3717, 3720
- median residual time, 3758
- minimum time, 3717
- missing stratum values, 3721, 3730, 3734
- missing values, 3735
- modified Peto-Peto test for homogeneity, 3704, 3734
- Nelson-Aalen estimates, 3721
- ODS graph names, 3764
- ODS graphics, 3715
- ODS table names, 3762
- output data sets, 3753
- partial listing, 3728
- Peto-Peto test for homogeneity, 3704, 3734
- probability density function, 3704, 3790
- product-limit estimates, 3704, 3706, 3720, 3736, 3755–3757, 3766
- Scheffe's adjustment, 3731
- Sidak's adjustment, 3732
- simulated adjustment, 3732
- stratified tests, 3704, 3705, 3712, 3714, 3733, 3748, 3760
- studentized maximum modulus adjustment, 3732
- survival distribution function, 3704, 3736
- Tarone-Ware test for homogeneity, 3704, 3734
- traditional graphics, 3715
- transformations for confidence intervals, 3718
- trend tests, 3704, 3734, 3750, 3760
- Tukey's adjustment, 3732
- Wilcoxon test for association, 3705, 3750
- Wilcoxon test for homogeneity, 3704, 3734, 3746
- likelihood ratio test for homogeneity
  - LIFETEST procedure, 3704
- line printer plots
  - LIFETEST procedure, 3715
- linear rank tests, *see* association tests
- linear transformation
  - confidence intervals (LIFETEST), 3718, 3741
- local annotate
  - traditional graphics (LIFETEST), 3720
- log transformation
  - confidence intervals (LIFETEST), 3718, 3741, 3743
- log-log transformation
  - confidence intervals (LIFETEST), 3718, 3741, 3743
- log-rank test for association
  - LIFETEST procedure, 3705
- log-rank test for homogeneity
  - LIFETEST procedure, 3704, 3734, 3746
- logit transformation
  - confidence intervals (LIFETEST), 3718, 3742, 3743
- maximum time
  - confidence bands (LIFETEST), 3717
  - plots (LIFETEST), 3720
- mean survival time
  - time limit (LIFETEST), 3728
- median residual time
  - LIFETEST procedure, 3758
- minimum time
  - confidence bands (LIFETEST), 3717
- missing stratum values
  - LIFETEST procedure, 3721, 3730, 3734
- missing values
  - LIFETEST procedure, 3735
- modified Peto-Peto test for homogeneity
  - LIFETEST procedure, 3704, 3734
- multiplicity adjustment
  - Bonferroni (LIFETEST), 3731
  - Dunnnett (LIFETEST), 3731
  - Scheffe (LIFETEST), 3731
  - Sidak (LIFETEST), 3732
  - simulated (LIFETEST), 3732
  - studentized maximum modulus (LIFETEST), 3732
  - Tukey (LIFETEST), 3732
- Nelson-Aalen estimates
  - LIFETEST procedure, 3721
- number of intervals
  - life-table estimates (LIFETEST), 3721
- ODS graph names
  - LIFETEST procedure, 3764
- ODS graphics
  - LIFETEST procedure, 3715
- output data sets
  - LIFETEST procedure, 3753

- partial listing
  - product-limit estimate (LIFETEST), 3728
- PDF, *see* probability density function
- Peto-Peto test for homogeneity
  - LIFETEST procedure, 3704, 3734
- Peto-Peto-Prentice, *see* Peto-Peto test for homogeneity
- probability density function
  - LIFETEST procedure, 3704, 3790
- product-limit estimates
  - LIFETEST procedure, 3704, 3706, 3736, 3755–3757
- Scheffe’s adjustment
  - LIFETEST procedure, 3731
- SDF, *see* survival distribution function
- Sidak’s adjustment
  - LIFETEST procedure, 3732
- simulated adjustment
  - LIFETEST procedure, 3732
- stratified tests
  - LIFETEST procedure, 3704, 3705, 3712, 3714, 3733, 3748, 3760
- studentized maximum modulus adjustment
  - LIFETEST procedure, 3732
- survival distribution function
  - LIFETEST procedure, 3704, 3736, 3754
- survivor function, *see* survival distribution function
- Tarone-Ware test for homogeneity
  - LIFETEST procedure, 3704, 3734
- traditional graphics
  - LIFETEST procedure, 3715
- transformations for confidence intervals
  - LIFETEST procedure, 3718
- trend tests
  - LIFETEST procedure, 3704, 3734, 3750, 3760
- Tukey’s adjustment
  - LIFETEST procedure, 3732
- Wilcoxon test for association
  - LIFETEST procedure, 3705
- Wilcoxon test for homogeneity
  - LIFETEST procedure, 3704, 3734, 3746



# Syntax Index

ADJUST= option  
    STRATA statement (LIFETEST), 3731  
ALPHA= option  
    PROC LIFETEST statement, 3716  
ALPHAQT= option  
    PROC LIFETEST statement, 3717  
ANNOTATE= option  
    PROC LIFETEST statement, 3717  
ATRISK option  
    PROC LIFETEST statement, 3717  
  
BANDMAX= option, *see* BANDMAXTIME=  
    option  
BANDMAXTIME= option  
    PROC LIFETEST statement, 3717  
BANDMIN= option, *see* BANDMINTIME=  
    option  
BANDMINTIME= option  
    PROC LIFETEST statement, 3717  
BY statement  
    LIFETEST procedure, 3728  
  
CENSOREDSDYMBOL= option  
    PROC LIFETEST statement, 3717  
CONFBAND= option  
    PROC LIFETEST statement, 3717  
CONFTYPE= option  
    PROC LIFETEST statement, 3718  
  
DATA= option  
    PROC LIFETEST statement, 3718  
DESCRIPTION= option  
    PROC LIFETEST statement, 3718  
DIFF= option  
    STRATA statement (LIFETEST), 3733  
  
EVENTSYMBOL= option  
    PROC LIFETEST statement, 3719  
  
FORMCHAR= option  
    PROC LIFETEST statement, 3719  
FREQ statement  
    LIFETEST procedure, 3729  
  
GOUT= option  
    PROC LIFETEST statement, 3719  
GROUP= option  
    STRATA statement (LIFETEST), 3733  
  
ID statement

    LIFETEST procedure, 3729  
INTERVALS= option  
    PROC LIFETEST statement, 3719  
  
LANNOTATE= option  
    PROC LIFETEST statement, 3720  
LIFETEST procedure, 3703  
    BY statement, 3728  
    FREQ statement, 3729  
    ID statement, 3729  
    PROC LIFETEST statement, 3714  
    STRATA statement, 3730  
    syntax, 3714  
    TEST statement, 3735  
    TIME statement, 3735  
LIFETEST procedure, BY statement, 3728  
LIFETEST procedure, FREQ statement, 3729  
    NOTRUNCATE option, 3729  
LIFETEST procedure, ID statement, 3729  
LIFETEST procedure, PROC LIFETEST  
    statement, 3714  
    ALPHA= option, 3716  
    ALPHAQT= option, 3717  
    ANNOTATE= option, 3717  
    ATRISK option, 3717  
    BANDMAXTIME= option, 3717  
    BANDMINTIME= option, 3717  
    CENSOREDSDYMBOL= option, 3717  
    CONFBAND= option, 3717  
    CONFTYPE= option, 3718  
    DATA= option, 3718  
    DESCRIPTION= option, 3718  
    EVENTSYMBOL= option, 3719  
    FORMCHAR= option, 3719  
    GOUT= option, 3719  
    INTERVALS= option, 3719  
    LANNOTATE= option, 3720  
    LINEPRINTER option, 3720  
    MAXTIME= option, 3720  
    METHOD= option, 3720  
    MISSING option, 3721  
    NELSON option, 3721  
    NINTERVAL= option, 3721  
    NOCENSLOT option, 3721  
    NOLEFT option, 3721  
    NOPRINT option, 3721  
    NOTABLE option, 3721  
    OUTSURV= option, 3721

- OUTTEST= option, 3721
- PLOTS= option, 3722, 3726
- REDUCEOUT option, 3727
- SINGULAR= option, 3727
- STDERR option, 3727
- TIMELIM= option, 3728
- TIMELIST= option, 3728
- WIDTH= option, 3728
- LIFETEST procedure, STRATA statement, 3730
  - ADJUST= option, 3731
  - DIFF= option, 3733
  - GROUP= option, 3733
  - MISSING option, 3734
  - NODETAIL option, 3734
  - NOTEST option, 3734
  - TEST= option, 3734
  - TREND option, 3734
- LIFETEST procedure, TEST statement, 3735
- LIFETEST procedure, TIME statement, 3735
- LINEPRINTER option
  - PROC LIFETEST statement, 3720
- MAXTIME= option
  - PROC LIFETEST statement, 3720
- METHOD= option
  - PROC LIFETEST statement, 3720
- MISSING option
  - PROC LIFETEST statement, 3721
  - STRATA statement (LIFETEST), 3734
- NELSON option
  - PROC LIFETEST statement, 3721
- NINTERVAL= option
  - PROC LIFETEST statement, 3721
- NOCENSLOT option
  - PROC LIFETEST statement, 3721
- NODETAIL option
  - STRATA statement (LIFETEST), 3734
- NOLEFT option
  - PROC LIFETEST statement, 3721
- NOPRINT option
  - PROC LIFETEST statement, 3721
- NOTABLE option
  - PROC LIFETEST statement, 3721
- NOTEST option
  - STRATA statement (LIFETEST), 3734
- NOTRUNCATE option
  - FREQ statement, 3729
- OUTSURV= option
  - PROC LIFETEST statement, 3721
- OUTTEST= option
  - PROC LIFETEST statement, 3721
- PLOTS= option
  - PROC LIFETEST statement, 3722, 3726
- PROC LIFETEST statement
  - LIFETEST procedure, 3714
- REDUCEOUT option
  - PROC LIFETEST statement, 3727
- SINGULAR= option
  - PROC LIFETEST statement, 3727
- STDERR option
  - PROC LIFETEST statement, 3727
- STRATA statement
  - LIFETEST procedure, 3730
- TEST statement
  - LIFETEST procedure, 3735
- TEST= option
  - STRATA statement (LIFETEST), 3734
- TIME statement
  - LIFETEST procedure, 3735
- TIMELIM= option
  - PROC LIFETEST statement, 3728
- TIMELIST= option
  - PROC LIFETEST statement, 3728
- TREND option
  - STRATA statement (LIFETEST), 3734
- WIDTH= option
  - PROC LIFETEST statement, 3728

## Your Turn

---

We welcome your feedback.

- If you have comments about this book, please send them to **`yourturn@sas.com`**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **`suggest@sas.com`**.





# SAS® Publishing Delivers!

Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at [support.sas.com/bookstore](http://support.sas.com/bookstore).

## SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

**[support.sas.com/saspress](http://support.sas.com/saspress)**

## SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – **free** on the Web.
- Hard-copy books.

**[support.sas.com/publishing](http://support.sas.com/publishing)**

## SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

**[support.sas.com/spn](http://support.sas.com/spn)**



**THE  
POWER  
TO KNOW®**

